

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:45:19 ; Search time 18 seconds
(without alignments)
1982.959 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSPFNRQWTSLSLRVTA.....LSVEEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3927	100.0	759	1 EPLI_HUMAN	Q9ubh6 homo sapien
2	2983	76.0	753	1 EPLI_MOUSE	Q9ergo mus musculu
3	239	6.1	915	1 NFM_HUMAN	P07197 homo sapien
4	231.5	5.9	704	1 NP14_RAT	P41777 rattus norv
5	222	5.7	2476	1 ATRX_MOUSE	Q61687 mus musculu
6	217	5.5	219	1 TSF3_HUMAN	P29675 helianthus
7	216	5.5	810	1 NFM_BOVIN	O77788 bos taurus
8	214	5.4	1233	1 YE16_YEAST	P43597 saccharomyc
9	213	5.4	6632	1 UN89_CABEL	Q01761 caenorhabdi
10	211	5.4	1790	1 US01_YEAST	P25386 saccharomyc
11	206	5.2	1961	1 MYH9_RAT	Q62812 rattus norv
12	205	5.2	848	1 NFM_MOUSE	P08553 mus musculu
13	201	5.1	1658	1 YN67_YEAST	Q03661 saccharomyc
14	201	5.1	2116	1 MYS2_DICDI	P08799 dictyosteli
15	200.5	5.1	1960	1 MYH9_HUMAN	P35579 homo sapien
16	200	5.1	699	1 NP14_HUMAN	Q14978 homo sapien
17	200	5.1	2845	1 APC_MOUSE	Q61315 mus musculu
18	199	5.1	845	1 NFM_RAT	P22839 rattus norv
19	196.5	5.0	1938	1 MYS_AEQU	P24733 aequipecten
20	196	5.0	2492	1 ATRX_HUMAN	P46100 homo sapien
21	195.5	5.0	1427	1 REST_HUMAN	P30622 homo sapien
22	194	4.9	5120	1 PCLO_CHICK	Q9pu36 gallus gall
23	193	4.9	2843	1 APC_HUMAN	P25054 homo sapien
24	193	4.9	3866	1 HRX_MOUSE	P55200 mus musculu
25	192.5	4.9	1616	1 P200_MYCBE	Q49429 mycoplasma
26	192	4.9	1308	1 MAK6_MOUSE	Q9jms2 mus musculu
27	192	4.9	1453	1 NKCR_MOUSE	P30415 mus musculu
28	192	4.9	1781	1 AK12_HUMAN	Q02952 homo sapien
29	190	4.8	1020	1 NPH_HUMAN	P22036 homo sapien
30	190	4.8	1332	1 MAK5_HUMAN	Q9n4c8 homo sapien
31	188.5	4.8	960	1 YMK6_YEAST	Q04279 saccharomyc
32	187.5	4.8	3969	1 HRX_HUMAN	Q03164 homo sapien
33	186.5	4.7	644	1 NFM_RABIT	P54938 oryctolagus

34	185.5	4.7	944	1 NUFL_YEAST	P32380 saccharomyc
35	184.5	4.7	954	1 BIR1_YEAST	P47134 saccharomyc
36	184	4.7	1341	1 ACIN_HUMAN	Q9ukv3 homo sapien
37	183.5	4.7	1939	1 MYH6_HUMAN	P13633 homo sapien
38	182.5	4.6	1395	1 SP41_YEAST	P38904 saccharomyc
39	181	4.6	5147	1 PCLO_HUMAN	Q9v6v0 homo sapien
40	180	4.6	1935	1 MYH7_PIG	P79293 sus scrofa
41	180	4.6	2774	1 MAPA_RAT	P34926 rattus norv
42	179.5	4.6	678	1 GARP_PLAUF	P13816 plasmodium
43	179.5	4.6	1530	1 SCP2_HUMAN	Q9bx26 homo sapien
44	179.5	4.6	1976	1 MYH4_HUMAN	P35580 homo sapien
45	179.5	4.6	2779	1 LVA_DROME	Q8msl1 drosophila

ALIGNMENTS

RESULT 1

EPLI_HUMAN

ID AC Q9UBH6; Q9BVF2; Q9H8J1; Q9HBN5; Q9NXC3; Q9NXU6; Q9P0H6; Q9UBH5;

AC Q9UBH5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Epithelial protein lost in neoplasm.

GN EPLIN OR SREBP3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

[1]

RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).

RC TISSUE=Cervical carcinoma;

RX MEDLINE=20087188; PubMed=10618726;

RA Maul R.S., Chang D.D.;

RT "EPLIN, epithelial protein lost in neoplasm.";

RL **Oncogene 18:7838-7841(1999)**

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).

RX MEDLINE=20267849; PubMed=10806352;

R Chen S., Maul R.S., Kim H.R., Chang D.D.;

RT "Characterization of the human EPLIN (Epithelial protein lost in neoplasm) gene reveals distinct promoters for the two EPLIN isoforms.";

RL **Gene 248:69-76(2000)**

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM BETA).

RC TISSUE=Uterus;

RX MEDLINE=21154917; PubMed=11230166;

RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H., Lauber J., Duesterhoft A., Beyer A., Koehrer K., Strack N., Mewes H.-W., Oesterwaelder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M., Poustka A.;

RT "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs.";

RL Genome Res. 11:422-435(2001).

[4]

RP SEQUENCE FROM N.A. (ISOFORMS ALPHA, BETA AND 3).

RC TISSUE=Colon, Hepatoma, and Placenta;

R Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RN Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.

[5]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Hypothalamus;

R Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X., Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H., Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,


```
QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFOST 600
Db 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFOST 600
QY 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERQVENAKASKKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERQVENAKASKKNGVNGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLENNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVDNTFAFEFT 720
Db 661 ETGKRSKEGHSLENNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVDNTFAFEFT 720
QY 721 TONOKSQDVELWEGEVVKELSVBEQIKRNYDEDEDEE 759
Db 721 TONOKSQDVELWEGEVVKELSVBEQIKRNYDEDEDEE 759
```

RESULT 2

```
ID EPLI MOUSE STANDARD; PRT: 753 AA.
AC Q9ERG0; Q9ERG1;
JT 28-FEB-2003 (Rel. 41, Created)
JT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Epithelial protein lost in neoplasm (mEPLIN).
ZN EPLIN.
DS Mus musculus (Mouse).
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
DC NCBI_TaxID=10090;
ZN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RX MEDLINE=21100452; PubMed=1179679;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
RT and comparison of mammalian and zebrafish EPLIN.";
RL Gene 262.155-160(2001).
CC -! SUBCELLULAR LOCATION: Cytoplasmic. This cytoskeletal protein co-
CC localizes with actin stress fibers.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Beta;
CC IsoId=Q9ERG0-1; Sequence=displayed;
CC Name=Alpha;
CC IsoId=Q9ERG0-2; Sequence=VSP_003118;
CC -! TISSUE SPECIFICITY: Eplin-alpha is highly expressed in embryos
CC from day 7-11 and in adult spleen and lung. Eplin-beta expression
CC is highest in adult kidney, testis, lung and liver, intermediate
CC in heart, brain, spleen, skeletal muscle and low in embryos.
CC -! SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
RR EMBL; AF307844; AAC31147.1;
RR EMBL; AF307845; AAC31148.1;
RR HSSP; P32965; ICTL.
RR MG; MGI:192092; Eplin.
RR GO; GO:0015629; C:actin cytoskeleton; IDA.
RR InterPro; IPR001781; LIM.
RR Pfam; PF00412; LIM; 1.
RR ProDom; PD000094; LIM; 1.
RR SMART; SM00132; LIM; 1.
RR R PROSITE; PS00478; LIM DOMAIN 1; 1.
RR R PROSITE; PS50023; LIM DOMAIN 2; 1.
RR W Cytoskeleton; LIM domain; Metal-binding; Zinc; Alternative splicing.
RR DOMAIN 386 446 LIM.
RR VARSPLOC 1 160 Missing (in isoform Alpha).
```

```
FT 216 216 /FTID=VSP_003118.
FT 325 325 S -> N (IN REF. 1; AAC31148).
FT 486 486 A -> T (IN REF. 1; AAC31148).
FT 499 499 P -> S (IN REF. 1; AAC31147).
FT 538 538 V -> A (IN REF. 1; AAC31147).
FT 538 538 S -> G (IN REF. 1; AAC31148).
FT 605 605 T -> A (IN REF. 1; AAC31148).
SQ SEQUENCE 753 AA; 84089 MW; 9A4692E86DF4A2AF CRC64;

Query Match 76.0%; Score 2983; DB 1; Length 753;
Best Local Similarity 76.3%; Pred. No. 7e-134;
Matches 580; Conservative 60; Mismatches 112; Indels 8; Gaps 3;

QY 1 MESSFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEETNMKKRSNTENL 60
Db 1 MESTFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEETNMKKRSNTENL 60
QY 61 SQHFRKGTILTVLKKWENPCLGAESHTLSLRNSSTEIRHRDHPHAEVTSHTAASAKADQ 120
Db 61 PQHFRGTLSVLKKWENPVAGAEFTDLSLPNSSSEGHTADYPPAEVTDKPAQGVADR 120
QY 121 EEOIHPRSRLRSPPEALVQGRYPHIKDGEDLDKDHSTESKMCNCLGSGRHEVEKSEISEN 180
Db 121 EEOIHPRSRLRSPPEALVQGRYPHIKDGEDLDKDHSTESKMCNCLGSGRHEVEKSEISEN 180
QY 121 EHTQPKPRGSRPEAVIQSRIYRSENHSHDFKAQATESQKMCNCLGSGRHEVEKSEISEN 180
Db 121 EHTQPKPRGSRPEAVIQSRIYRSENHSHDFKAQATESQKMCNCLGSGRHEVEKSEISEN 180
QY 181 TDASGKIEKYNVPLNRLKMWFEKGEPTQTKILRAQSRGASGRKISSENSYSLDDLEIGPQ 240
Db 181 TTSCKIEKYNVPLNRLKMWFEKGEHTQTKSLWTSRAGGRRLSENCSLDDWIGAGH 240
QY 241 LSSSTFDSEKNSRRNLPLRSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGEIKI 300
Db 241 LSSSAFNSKESKRNLEPLRSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGEIKI 300
QY 301 HWEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDSDRSQSVKSEVQVPHKPK 360
Db 299 HWEQKENVPPGPEACSVHQEGSVATENSLVALSVPAEDDTCNSQVKEAQQPMHKKP 358
QY 361 LSPDRASSLSSESPKAMKFKQAPARETCVCEQKTVYPMERLLANQOVFHTSCRCVC 420
Db 359 LSPDARTSSLPESSPKTAKKQAPAKSCVCEQKTVYPMERLLANQOVFHTSCRCVC 418
QY 421 NNKLSLGTYSALHGRIYCKPHFNLFKSGNYDEGFHGRPHKDLWASKNENEELERPAQ 480
Db 419 NNKLSLGTYSALHGRIYCKPHFNLFKSGNYDEGFHGRPHKDLWASKNENEELERPAQ 478
QY 481 LANARETHSPGVEDAPIAKYCVLAASMEAKASQOEKEDKPAETKRLRIAWPPPTLGS 540
Db 479 PPNAGESPHSPGVEDAPIAKYCVLAASMEAKASQOEKEDKPAETKRLRIAWPPPTLGS 538
QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFOST 600
Db 539 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFOST 598
QY 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERQVENAKASKKNGVNGKTTWQNKESK 659
Db 599 SVKSPKTVSPPIRKWMSQSESVGGRVAERQVENAKASKKNGVNGKTTWQNKESK 653
QY 660 ETGKRSKEGHSLENNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVDNTFAFEFT 719
Db 654 GEEVPRSKDRSSFELESENFENANGAETDNDHVAQSPLEPAQPGHSGFVDITAAKEF 713
QY 720 TONOKSQDVELWEGEVVKELSVBEQIKRNYDEDEDEE 759
Db 714 TONOKSQDVELWEGEVVKELSVBEQIKRNYDEDEDEE 753
```

RESULT 3

```
NFW_HUMAN
ID NFW_HUMAN
AC P07197;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M) (Neurofilament 3).
GN NEF3 OR NEFM OR NFM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini, Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87275853; PubMed=3608989;
RA Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;
RT "The human mid-size neurofilament subunit: a repeated protein
RT sequence and the relationship of its gene to the intermediate
RT filament gene family";
RL EMBO J. 6:1617-1626(1987).
RN [2]
RP SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.
RX MEDLINE=88158120; PubMed=2450354;
RA Lee V.M.-Y., Otvos L. Jr., Carden M.J., Hollosi M., Dietzschold B.,
RA Lazzarini R.A.;
RT "Identification of the major multiphosphorylation site in mammalian
RT neurofilaments";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Y00067; CAA68276.1; --
CC PIR; A27864; A27864
CC Genew; HGNC:7734; NEF3.
CC MIM; 162250; --
CC GO; GO:0005883; C:neurofilament; TAS.
CC GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
CC InterPro; IPR006821; Filament_head.
CC InterPro; IPR001864; IF.
CC InterPro; IPR002957; Keratin_I.
CC Pfam; PF00038; filament; 1.
CC Pfam; PF04732; filament_head; 1.
CC PRINTS; PRO1248; TYPE1KERATIN.
CC PROSITE; PS00226; IF; 1
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 103 HEAD.
FT DOMAIN 104 411 ROD.
FT DOMAIN 412 915 TAIL.
FT DOMAIN 104 135 COIL 1A.
FT DOMAIN 136 148 LINKER 1.
FT DOMAIN 149 247 COIL 1B.
FT DOMAIN 248 264 LINKER 12.
FT DOMAIN 265 286 COIL 2A.
FT DOMAIN 287 290 LINKER 2.
FT DOMAIN 291 411 COIL 2B.
FT DOMAIN 613 690 5 X 13 AA TANDEM REPEATS.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).

SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DF34C9D9E50 CRC64;
Query Match 6.1%; Score 239; DB 1; Length 915;
Best Local Similarity 22.1%; Pred. No. 0.0002;
Matches 180; Conservative 117; Mismatches 318; Indels 198; Gaps 39;
QY 15 SLRVTALEXSLVNNKSSAIVEIFSYQKAAEETNMKKESNTE----NLQSHFRKGTIT 70
DB 204 ALRKDIEEASLVK-----VELDKKVSQSLQDEVAF--LRSNHEEEVADLLAQIASHIT 254
QY 71 VLKKWENPGLG-----ESHTDSLNSSTE---IHRADHPAETVSHAASGAK- 117
DB 255 VERKDYLTDTISTALKEIRSQLSHSDQNNHQAEWFKCRYAKLTEAAEQNKKAIRSAKE 314
QY 118 --ADQEEQIHPRS-----RLRSPPEAL-----VQGRYPHIKDGEDLKDHSSTKKWENCL 165
DB 315 EIAEYRRQLQSKSIELESVRGTESLERQLSDLEERHNN-----DLSSYQDTTQOENEL 369
QY 166 GSRHEVEKSEISENTDASGKIEKYNVPLNRLKMFKEKGEPTQTKILRAQSRSARGKIS 225
DB 370 RGTKEWAR-----HLEFYQDLLN-VKMAIDIEIAAYRKLEEGE-----TRFS 412
QY 226 ENSYSLDLEIGFQQLSSSTFDEKSNESRNLELPEL-----SETSIKRMAYK 274
DB 413 TFAGSI-----TGPLYTHRPDITTSKIQTKEAPKLVQHKFVEIIBETKVEDEKSEM 468
QY 275 QAAVSKOSSSTNTNELKASGGIKIHQMEQKNVP-----PCPEVCIT 318
DB 469 EEAL-----TAITEELAAKMEKEKAAEKEEPEEABEEVAAKKSPYKATAPEV--K 520
QY 319 HQGEK-----ISAMENSLAVRSTPAEDSDRSOVKSEVQOPVHPKPLSPDSRASLSLSE 372
DB 521 EBSGEKEEBSGEKEEEDGAKSDQAEEGSGEKEGSEKEGEQEE--GETEAEAGE 577
QY 373 SSPPKAMKFPAPARTCEQKTVPMERLLANQQVFPHISCFRCSCYCNKSLGTVASL 432
DB 578 EAAKEEKKVEEKSEEVATK-----EELVADAKVE-----607
QY 433 HGRIYCKPHNPQFKSGNVDGFGHRPHKOLWASKNENEIEILERPAQLANARETPHSPG 492
DB 608 -----KPEKAKGPVKSPVEEK-GKSP-----VPKSPVEEKGPVKSPVEEKGKSP- 654
QY 493 VEDAPITAKVG---VLAASMEAKASSQOEKEDPAETKURIAHPPPTTELGSSSALSEGI 549
DB 655 VPKSPVEEKGPVKSPVEEKAKSPVK--SPVEEAKSK-----AEVG-KGEQKEEE 705
QY 550 KMSKPKWPPDEI-SKPEVPEDVDLDLKLRRSSSLKER--SRPFTVAASFQTSVUSPK 606
DB 706 KEVK-EAPKEEKVEKEEKPKVP---EKKAESPVKEEAFAEVVTTITKSVKVLHSEKTK 761
QY 607 TVSPPIRKGMSESESESESVGG--VAERKQVENAKASKK-----NGNVGKTTWQNKESK 659
DB 762 EEGKPLQ-----EKEKAKAGGEGGSEESGSKGAKGSRKEDIANGEV-----EGK 808
QY 660 GETGKRSKEGHSLEMENENLVENGAD---SDEDDNSFLKQQSPQEPKPSLNSWFSVNTFA 716
DB 809 EYVEQETKEGSGREBEKGVYTNGLDLSPADEKKGDKSEKVVVTKTEKITSEGGDGA 868
QY 717 EEFITTON---QKSQD-VELWEGEVVKELSVEE 744
DB 869 TKYITKSVTVTKVBEHEETFEKLVSTKKVEK 901
RESULT 4
NP14 RAT ID NP14 RAT STANDARD; PRT; 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE phosphoprotein 1).

Matches 161; Conservative 112; Mismatches 288; Indels 227; Gaps 311

```

QY 30 KSSAIVEIFSKYOKAAAEETNMEKKRNTENLSOHFRKGLTIVLKKWENPGLCAE-----84
D 23 RDNLQSEVASKFAKATCATQODANASSLLDIYFWLKST-KAPKVKLQSGNPVAKKAKKE 81
QY 85 -SHTDSLNRNSSTIIRHADHP-----AEVTSHAASGA-----116
D 82 TSSDSDASESSEB-EKAQVPTOKAAPAKRASLPQHAGKAAAKAKASESSSESSSEEE 140
QY 117 -----KADQEQIHPRSR-LRSPPEALVOGRYPHIKDGEDLKDHSSTESKMWENCLGESRH 170
D 141 KDKKKXPVQOKAVKPPPK-----KAESSESDSSSEDEAPQTKPKAAA 192
QY 171 EVKSEISENTDASGKIEKYNVPLNKLWMFEKGEPTQKILRAQSRASGRKISSENSYS 230
D 193 TAAKAPTKAQTKAPAK-----PGPPA-----KAQPKAANGKAGSSSS-- 229
QY 231 LDLELTGPGOLSSSTFDSSEKNESRRNLELPRISETSIKORM-----AKYQAAVSKOSSST 285
D 230 -----SSSSSSDDSEBEKKAAP-LKKTAPKKVQVAKAPVKVTAAPTQKSSSS 277
QY 286 NYTNELKASGGEIKIHKEKE-----NVPPGPEVCIITHQEGEKISANENSIAVSTPA 339
D 278 E-----DSSSEBEEQKPKMKKAGYSSVPP-PSVLSK---KSVGASPPKAAAQTOP 328
QY 340 EDDSRDSQVKSEVOQVHPKPI.SPDGRASLSBSPKAMKFOAPARETCVECKTVYP 399
D 329 ADSASDS-----SEEDSSSEBEKKTPAKTVVSKTPAKPAPVK-----366
QY 400 MERLLANQVFIHSCPRCSYCNKLSLGTYSALHGRFYCKPHNQLFKSGNYDEGFGR 459
D 367 -----KKAESSSDSS-----DSDSSEDEA----385
QY 460 PHKDLWASKNENEIEILERA-----OLAVARTPHSP-GVEDAPIAKGVGLAASMEAKAS 513
D 386 PAKPVSAATKSP-----LSKPAVTPKPPAAKAVATPKQAPAGSGQKPKQSKADSSSBESSS 441
QY 514 SQQEKEDKPAETKRLI-----AWPPPTLSSGSGALBEGIKMSPKPPPEDEISK 564
D 442 SEBEATKSVTTPKARVATAKAPSLFQAKQAPRAGDSSDSSSESSSEBEKKTTP-----K 496
QY 565 PEVPEDVDLKLRRSSSLKERSRPTVAASFQSTSVKS PKTVPPIKGMKSMSEQSEE 624
D 497 PPA-----KKGAAGAAMPKPTPKKAAAESSSSSSSDSSSEBEKKK-PKSKATPK 546
QY 625 SVGGRVAERKQVENAKASKKNGVKTWTWQNKESKG-ETGKRKSEGHLSMELENENLVENG 683
D 547 PQAGKANGVPASQNGKAGSESEEEEDTQNKAAAGTKPGSGKKRKH-----ETA 597
QY 684 ADSDEDNDFLKKQSQPEKPSLNWSSFYDNTFAEEFTTQNKSQDVVELWEGEVVKLSVE 743
D 598 DEAAATPQSKVKVLQTP-----NTTPKPKKGEKRASSPPRRVREE---EIEYD 641
QY 744 EQIKRNRY 751
D 642 SRVADNSF 649

```

RESULT 5

ATRX_MOUSE

ID	ATRX_MOUSE	STANDARD	PRT	2476 AA.
AC	G61687;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Transcriptional regulator ATRX (X-linked nuclear protein)			
DE	(Heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38 protein)			
DE	GN			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			

DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Pollen specific protein SF3.
 GN Helianthus annuus (Common sunflower).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
 OC Heliantheae; Helianthus.
 OX NCBI_TaxID=4232;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, HA401B / Cargill; TISSUE=Pollen;
 RX MEDLINE=93258417; PubMed=1302629;
 RA Baltz R., Dorn C., Pillay D.T.N., Steinmetz A.;
 RT "Characterization of a pollen-specific cDNA from sunflower encoding a
 zinc finger protein.";
 RL Plant J. 2:713-721(1992).
 RN [2]
 RP IDENTIFICATION OF LIM DOMAINS.
 RX MEDLINE=93104672; PubMed=1467648;
 RA Baltz R., Eward J.-L., Dorn C., Steinmetz A.;
 RT "A LIM motif is present in a pollen-specific protein.";
 RL Plant Cell 4:1465-1466(1992).
 CC -!- FUNCTION: COULD POSSIBLY INVOLVED IN CONTROLLING POLLEN-SPECIFIC
 PROCESSES SUCH AS MALE GAMETE MATURATION, POLLEN TUBE FORMATION,
 OR EVEN FERTILIZATION.
 CC -!- TISSUE SPECIFICITY: POLLEN.
 CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; X64392; CAA45731.1; -
 DR PIR; S28507; S28507.
 DR HSSP; P04006; LIML.
 DR InterPro; IPR001781; LIM.
 DR Pfam; PF00412; LIM; 2.
 DR ProDom; PD000094; LIM; 2.
 DR SMART; SM00132; LIM; 2.
 DR PROSITE; PS00478; LIM DOMAIN 1; 1.
 DR PROSITE; PS0023; LIM DOMAIN 2; 2.
 KW Repeat; LIM domain; Metal-binding; Zinc.
 FT DOMAIN 9 109 LIM 1.
 FT DOMAIN 110 167 LIM 2.
 FT SEQUENCE 219 AA; 24832 MW; A74BDD1BB1393DE6 CRC64;
 Query Match 5.5%; Score 217; DB 1; Length 219;
 Best Local Similarity 35.8%; Pred. No. 0.0004;
 Matches 48; Conservative 23; Mismatches 45; Indels 18; Gaps 4;
 QY 379 MKKQAPARETCVEQKTYVPMERLANQVPHISCFRCSCYNNKLSLGTVASLHGRYIC 438
 Db 1 MKSFTGTTQK-CTVCEKTYLVLDKLVANQRYVYHKACFRCHCNSTLKLNSNSFDGVYIC 59
 QY 439 KHFNQLKSKNGYDEGCHRPDKLWASKNENEILERPQLANARETPHSPGVEDA-- 496
 Db 60 RHFDQLFKRTGSLKSPDGP-----KPKPRTFSQTSAN-RUSSFFEGTRDKCN 111
 QY 497 -----PIAKGV 503
 Db 112 ACAKIVYPIRVKV 125

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE Neurofilament medium polypeptide (NF-M) (Fragment).
 GN NEF3 OR NEFM OR NFM.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Hill W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
 RA Gearhart D.A.;
 RT "The bovine neurofilament M subunit has a novel set of KSP repeats
 normally restricted to NF-H.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS
 PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 OF AXONAL CALIBER (BY SIMILARITY).
 CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
 SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; AF091342; AAC36357.1; -
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin I.
 DR Pfam; PF00038; filament; 1.
 DR PRINTS; PR01248; TYPEKERATIN.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT NON TER 1 296 ROD.
 FT DOMAIN 297 810 TAIL.
 FT DOMAIN <1 20 COIL 1A.
 FT DOMAIN 21 33 LINKER 1.
 FT DOMAIN 34 132 COIL 1B.
 FT DOMAIN 133 149 LINKER 12.
 FT DOMAIN 150 171 COIL 2A.
 FT DOMAIN 172 175 LINKER 2.
 FT DOMAIN 176 296 COIL 2B.
 FT DOMAIN 503 582 8 X 13 AA TANDEM REPEATS.
 FT SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;
 Query Match 5.5%; Score 216; DB 1; Length 810;
 Best Local Similarity 20.3%; Pred. No. 0.0021;
 Matches 170; Conservative 112; Mismatches 320; Indels 234; Gaps 37;
 QY 15 SLRVTAKELSLVNNKSSAIVEIFSYQKAAETNMKKRSNTE----NLSQHFRTGLT 70
 Db 89 ALRKDIIESSLVK-----VELDKKQVSLQDVAFA--LRSNHEEVEDLAIQIASHIT 139
 QY 71 VLKKKWNPCGLA-----ESHTDSLRSNSTE---IRHRADHPPEAVTSHAAGAK- 117
 Db 140 VERKDYLTIDISTALKEIRSQLSHSDQNMHQAEWFKCKYAKLTEAAEQNKAISSAKE 199
 QY 118 --ADQEQIHPRS----RLRSPEAL-----VQGRYPHIKQGEDLKHSTSKWENCL 165

```

Db 200 EIAEYRQLOSKSIELESVRGTHESLEROLSDIEERHNNH-----DLSSVQDTIOOLENEL 254
QY 166 GESRHEVEKSEISENTDASGKIEKYVNPVLRKMKPEKGEPTQTKILRAQSR--SASGRK 223
Db 255 RGTKEWAR-----HLREYQDLN-VKALDILIAAYRKLLEGEETRFPTFAGS 302
QY 224 ISENSYSLDLLEIGPQOLSSSTFDSKNSRRNLEPLRL-----SETSIKDRMA 272
Db 303 ITGPLYT---HRQPSIAISSKIQTK-----VEAPKLKVQHKKFVEBIEETKVEDEKS 352
QY 273 KYQAAVS--KQSSSTNYTHELKASGGEIKIHKMEQKENY-----PPGEVETITQHE 321
Db 353 EMBEALTAITEELAVSKVEKEEAEKEEKEEAEVEVVAAKGFVKATAPE--LKSEE 410
QY 322 GEK-----ISANENSLAVRSTPAEDSDSDOVKSEVQVPHPKPLSPDRASSLSSESP 375
Db 411 GEKEEREGOESEEBEAKSDQAEQSGGSEKEGESEGESEGESEGESEGESEGESEGESE 470
QY 376 PKAMKFFQAPARETCVECTVPMERLLANQOVPHISCFRCYCNKLSLGTYSALHGR 435
Db 471 AKEBKWEKAEB-----VAPKEELAAEAKV-----496
QY 436 IYCKPHFNQLFKSGNYDGFGRPHKDLWASKVNEEILERPQALANARETTPHSPGVE- 494
Db 497 -----EKPEK-----AKSPVAKSPTTKSPATAKSPAKSPEA 527
QY 495 -----DAPIAKVGVLAAAMEAKASSQOEKEDKPAETKRLRTAWPPP-----535
Db 528 KSPTAKSPTAKSPVAK-SPTAKSPEAKSPEAKSPTAKSPTAKSPTAKSPTAKSPTAKS 586
QY 536 -----TELSSGSALBEGIKMSKPKPPPEDEISKPEVPEDVDLDLKK-----LRRSSSLKE 586
Db 587 KABAGAEKGEQKEVEEKEAKESPKKEKAEKKEKPKDVP-EKKKABSPVKAASPVKE 645
QY 587 R--SRPFTVAASFQSTSVSKPTVSPPIKKGWSMSEQSEESVGRVAERKQVENAKASKK 644
Db 646 EVPAKPKVKVSPSE-----KEAKEEERPOEK--EKEKEKVEEVGG-----KEEGGLKESRK 692
QY 645 -----NGNVGKTTWQNKSGKETGSKSGHSLMELENLNVENGADSDDDNSFLKQKSP 699
Db 693 EDIAINGEV-----EGKEEETKEKSGSGE-----BEKGVTNGLDVSPGDE--KKGSD 740
QY 700 QEPKSLNWSFVDNTFAE-----BFTTQNKQSD-VELWEGEVVKLSVEE 744
Db 741 KSEKVVTKWKEKISSEGDDGATKVTIKSVTVTKVVEHEETFEKLSTKKVEK 796

RESULT 8
YF16 YEAST STANDARD; PRT; 1233 AA.
AC P43537;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 137.7 kDa protein in UGS1-PAB1 intergenic region.
GN YF016C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; D50617; BAA09255.1; --
DR PIR; S56271; S56271.
DR SGD; S0001912; YF016C.
KW Hypothetical protein.
SQ SEQUENCE 1233 AA; 137697 MW; CBA7CD2CGF0892F6 CRC64;

Query Match 5.4%; Score 214; DB 1; Length 1233;
Best Local Similarity 20.5%; Pred. No. 0.0042;
Matches 183; Conservative 125; Mismatches 308; Indels 278; Gaps 46;

QY 43 KAABETNKKRNS--TENLSQHFRKG-----TLTVLKKWENPGLGAESHTDS 89
Db 334 KDVSSESLTKGNFNFKENESKHLKAGEKQTESDRDGISPSVLAKNQKETEIGKDHVFE 393
QY 90 LRNSSTE-----IRHRADHPPAEVTSHAASGAKAD-----QEEQIHPRSRL 130
Db 394 QKQDEKCRKELSVNHEN-----MSHNFNAAGSDSIIPPETERETYDDETGMPTKRI 447
QY 131 RSPPEALVQR-----YPHIKDGEDLKDHSTESK-KMENCLGES-----RHEV---EKS 175
Db 448 SDNEKNLQHGNDISVEVEKEEEEEEEENSTFSVKYKENVTEGEQAVRNNEVSGTEEE 507
QY 176 EISENTDASGKIEKYVNPVLRKMKMFEKGEPTQTKILRAQSRGASGRKISSENSYLDLDE 235
Db 508 STSGKEEIMGDEKQS-----EAGE--KSSIIEIG-SANSAKISKDNLVLDEEA 554
QY 236 IGPQLSSST-----FD-----SEKNESRRNLELP-----RLSE--T 265
Db 555 EAPTQENKPTFWGGEIDIPADPRDDVIVEAVEKXIIIPEDLEVAKEQGEQGVKLDEPVK 614
QY 266 SIKRMAYQAAVS-----KQSSSTNYTNEL-----KASGGGIKIHMKQKE 307
Db 615 AMKDDKIAMRGAEISSEDMKKKQEGTAELSNKAKKEVDETARESAEGVEV-----EKS 668
QY 308 NVPPGPEVC--ITHQEGEKISANENSLAVRSTPAEDDSR--DSOVKSEV-----352
Db 669 KTPESPKVVRCTSGRPEDIQINERDPEV---LKEDVRVPDEDVKPEIATTIENSEED 724
QY 353 -----QQPVHPKPLSPDSRASSLSSESPPKAMKKFQAPARETCVQCQTVTPME 401
Db 725 PKSORVQISTEQAETTKQDMGDVGSITSFKEEEKPKRFEITQEGDKITGD-----775
QY 402 RLLANQOVFHISCFRCYCNKLS-LCTYASLHGRIYCKPHNQLPKSKGNVDEGFGRHP 460
Db 776 ---TNHE--HGEATEAASSENSKASDVGT---AEKYIEPSSSVKKD-----T 814
QY 461 HKDLWASKVNEEILERPQALANARETTPHSPGV-----EDAPI-----498
Db 815 EDAEVENSEKTEFIKVKAELENL-DAPKEAVTAEUNKENEDVEVDETEDEAEVENSEKT 873
QY 499 -----AKVGVLAAAMEAKASSQOEKEDKPAETKRLRIAWPPPTFELSGSSALKEGKMS 552
Db 874 EFKVKAEGLNLDAPKAEVTAELUNKENEDVEV-----AATSKEDIETKCS 919
QY 553 KPKWPP-----EDEISKPE---VPEDVDLDLKLRRSSSLK-----ERSRPFTVAAS 596
Db 920 EPABTPEEDGCTCEAEVSKDAEAVTKEDENMENSKI--AEALKDVTGDOEIDIDINISDE 977
QY 597 FQSTSVKSPKTVSPPIKKGWSMSEQSEESVGRVAERKQVENAKASKGNVGTQWQ- 655
Db 978 FQ-----RTVELP-----ELEKQIDKNKGDEKELEVEETEKETSLPDLVVEENI 1022
QY 656 KESKGETGKSKGEGHSLDEM-ENENLVENGADSDDDNSFLKQKSPQEPKSLN----WSSF 710
Db 1023 TEENKVEIKQEEVEVSLQDFNETETISKEAPNNDE--NGFEDQSTRENPKKASADDIFKDI 1080

```

QY 711 VDNTEBETTONKSOQVELWEGEVKLSVE-----EOKENRYVDEDED 757
DB 1081 LDET--NEFLEQKIVDSSEL--NALLSLDADKSTTTQTEQSKKN--DKPQD 1128

RESULT 9
ID UN89_CABEL STANDARD; PRT: 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR CO9D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
assembly, encodes a giant modular protein composed of Ig and signal
transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
assembly, encodes a giant modular protein composed of Ig and signal
transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
RN [3]
REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofibrillar
lattice assembly begins with positional cues laid down in the
basement membrane and muscle cell membrane. UNC-89 responds to
these signals, localizes, and then participates in assembling an
M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U33056; AAF00542.1; -;
CC EMBL; AF003131; AAB54132.2; -;
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig-C2.
CC InterPro; IPR003006; Ig-MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RCSD.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; Ig; 47.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF05177; RCSD; 5.

DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGc2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH; 2; 1.
DR PROSITE; PS50835; IG-Like; 49.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.

DR Pfam: PF04871; Usol p115 C; 1.
KW Transp: Protein Transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216B9FD4818 CRC64;

Query Match 5.4%; Score 211; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.009;
Matches 179; Conservative 128; Mismatches 296; Indels 234; Gaps 38;

QY 7 NRQWTSLSRLVTAKELSLVKNKSSAIVEIFSKYKAAEET--NMEKKRSNTENLSQHF 64
DB 1096 NLKNELETKLETSKALKVENEER-LAEKIQLEKEATETKQQLNSLRANLESLEKEH 1154
QY 65 RGTTLVLKKNENPGIGAE-----SHTSLRNSSTEIRHRDPPPAEVTSHAA 113
DB 1155 E--DLAAQLKYEEQIANKEQVNEISQINDEITSTQENESIKKNDELEGEV----- 1207
QY 114 SGAKAQEQIIPRSLRPPALVGRPHDKGDLKDHSTSKMNCIGESRHEVE 173
DB 1208 KAMKSTSEEQSNLK--KSEIDAL-----NLQIKELKKQNETNEASLLESIGSYE 1254
QY 174 KS--EISENTDASGTEKYNVPLNRLKMMFEXGEPTQTKILQAQSRASGRKISNSVL 231
DB 1255 SETVKIKELQDCNFKEK--EVESEEDKLKASEDKNXYLEOKES-----EKIKE--EL 1305
QY 232 D----DLEIGPGLSSSTFDSEKSENRRLLEPLRSETSIKDR-----MAKYQAAVSKQ 281
DB 1306 DAKTTELKIQLEKYNLSKAKESSE-----ELSLRKKTSSEKKAEOLEKLNKIQIK 1361
QY 282 SSGTNTYNELKASGGEI-----KIHKME-----QKENVPPGPEVCITHOGEKISAN 328
DB 1362 NOAFERKLLNEGSGTITQYSEKINTLEDELIRLQNEULKEAIDNTRSELEKVSLS 1421
QY 329 ENSL-----AVRSTPAE-----DDSRDSQVSEVQVPHKPLSPDRASSUSE 372
DB 1422 NDELLEEKQNTKSLQDEILSYNDKILTRNDEKLLSIEDRNKRD--ESLKEQLRAAQESK 1479
QY 373 SPPKAMKFOAPARETCVETQVYPMERLLANQOVPHISCFRCSCNNKLSLGTYSAL 432
DB 1480 AKVEGLKLEBESSKEKAELEKSKEMMKL-----ESTIESNETELKSSMET----- 1527
QY 433 HGRIVCKPHFNQIFKSKNGNYDEFGHRPHK--DLWASKNENE-EILERPAQLANARETPH 489
DB 1528 ----IRKSDKLEQSKGAEDIDKLNQLEKSDLSIRINESKDIELKSKL----- 1574
QY 490 SPQVEDAPIAKGVVLAAASWEAKASQOE--KEDKPAETKKLRIAWPPPTELGSSGSALE 547
DB 1575 -----RIEAKSGSELETVQQLNNAQEKIRI-----NAEEN 1605
QY 548 GIKMSKPKWPPDEISKVPVDPVDLDLK----KLRRSSSLKE--RSRPFVVAASFQSTS 601
DB 1606 TVLKSK-----LEDIERELKQKAEIKSNQOEKELTSLRKELEQELDSTQ 1651
QY 602 VKSPKTVSPPIRKGMWSMQSESVGGRVAERK-OVENAKASKK-----NGNVGR-T 651
DB 1652 QRAQK-----SEEE--RAEVKRFQVEXSQLEKAMLETKYNLUNKEQ 1694
QY 652 TWONKE-----SKGTGKRSKEGHSLEMENENLVENGADSDDEDDNSFLKQSQPOEP 702

DB 1695 AKRDEDTVKTTDSQROEIEKLAKELDNLKAENSRSEIDDLML----- 1746
QY 703 KSLNWSSEVDNTFAEFTTQKQSDVELWEVGVKLSVEEQIKRNRYYDEDEDE 759
DB 1747 -----LVTDLDEKNAKYR-----SKLKDLGVE--ISSDEEDDEDEDE 1782

RESULT 11
MYH9 RAT
ID MYH9 RAT STANDARD; PRT; 1961 AA.
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain, type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.; Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS, CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND CAPPING.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
DR EMBL; U31463; AAA74950.1; --
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; myosin_N; 1.
DR Pfam; PF01576; myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; G-actin-binding; Actin-binding; Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Query Match 5.2%; Score 206; DB 1; Length 1961;
Best Local Similarity 19.5%; Pred. No. 0.017;


```
Db 250 SHITVERKDYLTDTISTALKEIRSOLECHSDQNHQAEWFKRYAKLTAAEQNKEAR 309
Qy 115 GAK---ADQEQIHPRS-----RLRSPPEAL-----VQGRYPHIKQGEDLKQHSSTSKXM 161
Db 310 SAKETIAEYRQLOQSKSIELESVRGTQKESLERQISOIERHNH-----DLSSYQDTIQOL 364
Qy 162 ENCLGESRHEVSEISENTDASGKIEKYNVPLNRLKMKMEKEGPTQTKILRAQSSASG 221
Db 365 ENEURGTQKEMAR-----HUREYQDLN-VQMALDIBIAAYRKLLEGE-----407
Qy 222 RKISENSYSLDLEIGP-----GQLSSSTFDSEKNSRRNLLEPRL-----SETSI 267
Db 408 TRSTFSGSI---TGPLYTHRQPSVTISSKIQTK--VEAPKLKQVHKVEVEIIEETKV 461
Qy 268 KDMVAKYQAAVSKQSSSTNYTNELKASGGIKIHKMEQKENVP-----PGPEVCITHQ 320
Db 462 EDEKSEMEETL-----TATAEELAAAKEEK-EAAEKEEPEAEKSPVKSPEAKEEBE 514
Qy 321 EGKE-----ISANENSLAVRSTPAEDDSRDQVSEVQOPVHPKPLSPDPSRASSLSSES 373
Db 515 EGKEEBEEOEBEEDGKVDQAEEGSEKSEKDEGEQEBE-EGETPAEGEGEE 573
Qy 374 SPKAMKKFOAPARETCVEQKTYPMERLLANQOVFHISCFRCSCYCNKKLSLGTVYASLH 433
Db 574 AEAKEEKKIEGVEEVAVK-----592
Qy 434 GRIYKPHFNQLPKSGNYDEGFGHRPHKDLWAKNENEI-LERPAQANARETPHSPG 492
Db 593 -----BEIKVEKPEKAKS---PMPKSPV 612
Qy 493 VEDAPAKVGLAASMEAKASSQOEKEDKPAETKURIAMPPPTTELGGSSGSALEEGIKMS 552
Db 613 EYKVP-----KPEAKAGKEQKEEKEVEKESPE-----648
Qy 553 KPWPEDEISKPEVPEVDVLDLKLRRSSSLKERS--RPFTVAASFQSTVKSPTVSP 610
Db 649 -----EKVEKKEKPKDV---ADKKAEKSPVKEKAVEVITISKVKVSLKOTKEEKP 699
Qy 611 PIRKGSMSQSESVGRVAERKQVENAKASK-----NGNYGKTTWQNKESKGETGR 665
Db 700 QPOB--KVKEKAEEGGSE--EEGSDRSQESKEDIANGEV-----EGKEEBEOE 747
Qy 666 SKGHSLEMNENLVENGADSDDDNSFLKQSQPQPKSLNWSFVDNTFAEFTTQNOK 725
Db 748 TOEGSGREBEKGVTNGLD-----VSPAEEKXGDESS-----DDKVVVTK 789
Qy 726 SQDVELWEGE-----VVKELSVQEIKNRYDEDEDEE 759
Db 790 VEXITSEGGDGATKYITKSVTVTQKVEEH-----EETFE 824
```

RESULT 13

```
Ym67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUAI-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odeh C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.,
```

"The nucleotide sequence of Saccharomyces cerevisiae chromosome XIII.";

Nature 387:90-93(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z49809; CA489934.1; -

EMBL; Z49939; CA490190.1; -

PIR; S55101; S55101.

SGD; S0004832; ESC1.

GO; GO:0005634; C:nucleus; IDA

GO; GO:0006348; P:chromatin silencing at telomere; IMP.

Hypothetical protein.

SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match

Best Local Similarity 5.1%; Score 201; DB 1; Length 1658;

Matches 173; Conservative 146; Mismatches 312; Indels 320; Gaps 37;

Qy 24 SLWNKNSAIVEIPSKYKAAETNMKRSN-----TENLSQHFRKGLTLVL 72

Db 94 SLINRGKS--ILITLEK-EDALFERSLEERQRFQHLHSLMKNKYTGNSKSHQR---LIDL 147

Qy 73 KKKWENPGLAESHTDSLRNSSTEIRHRA--DHPPAEVTSAAAGAKADQEEIHPRL 130

Db 148 RK-----SOYGTDTSFQNNDEIPLDSFISSPLPDAEDESSNIDSKDEDEGKQSL 199

Qy 131 -----RSPPEALV-----QGRYPHIKGE-----DLK 152

Db 200 IKDFLENDEVELSEKNSQGSPPSIMLSDEYAEAGALQDVNSDEYAEEGQVPERK 259

Qy 153 DHSTESKKWENCLG-----ESRHEVEKSEISENTDASGK 186

Db 260 NIGQEQANVENATQISSSDSSEGGQYSEGVMELEDDIDVEDSAEKDESGAEGTEHVD 319

Qy 187 TEKNVPL--NRLKMMFEKGEPTQTKI-----LRAQSR 217

Db 320 FSKYMQPRTDKTPIVISKYSEDEHKVHQYSEDGAPFGSVNTSVDDSEDESEQAESY 379

Qy 218 SASGRKI--SENSYSLDLEIGFGQLSSSTFDSEKNSRRNLLEPRLSETSIKDRVAKYQA 276

Db 380 SANAENVYHNEHELDKEL-----IEDIESDSSESQSAQESQGSDDDFEYKM-KNEK 432

Qy 277 AVSKQSSSTNYTNELKASGGGIKHKMEOKEN--VPPGPEVCIT-----HQEKEKISAN 328

Db 433 STSEETENTSERDQGFADKAYTKKVEQOEENDEEPEKDDIIRSLDKNFGNNKSEYS 492

Qy 329 ENSLAVRSTPAEDSDRSQVSEVQOPVHPKPLSPDGRASSLSSESP-----PKAMKK 381

Db 493 ENVLNETDPAIVE-RENOIND-----VEGVDVTGKSVESDLHHSPLNLYDLAARMLQ 546

Qy 382 FOAPARETCVCEQKTYPMERLLANQOVFHISCFRCSCYCNKKLSLGTVYASLHRIYCKPH 441

Db 547 FQGRNSNCPQKEEQV-----SESYLCHSNGSNLSGR-----578

Qy 442 FNQLFKSGNYDEGFGHRPHKDLWAKNENIEILERPAQANARETPHSPGVEDAPIAKV 501

Db 579 -----SLDESEEQPLKDFTCENNNN-----LKTDR-----604

Qy 502 GVLAASMEAKASSQOEKEDKPAETKURIAWPPPTTELGGSSGSALEEGIKMS-----552

Db 605 GDLSSSVEIEVEKVESEKK-LDGSTEKELVPLSTDTTINNSLGNEDSIYSLDDADAISE 663

Qy 553 -----XPKW-----PPDEISKPEVPE---DVIDDLK 576

Db 664 NLTDVPLWEIKTTPKYEVVISESVYSTSVSTVEDNTVAMPPOVEYTPFPNDFNSLNDYE 723

QY 577 KLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRKWMSMSQSESVGRVAERKQV 636
 Db 724 K--RHLLKS-----TLAALAPATTKDAEFVEAGVTKSCLTSTSGHTNIFHTSKETQV 776
 QY 637 ENAKSKKNVNGVKTITWONKESKGETGRSKE-----GHSLEMEV----- 676
 Db 777 SLDLSTEN-----VTFENENTGDKNKQSKNFCQVANSTDKSTEDTNDKYFSAINYTN 831
 QY 677 -----ENLVENGADSD-----DMSFLKQSQPQPKSLNWSFVDNT 714
 Db 832 VTGSSCCDIETASNVENLRYCEKMNEMSSGDECVKQNDGSKTQISFSTQSPDN 891
 QY 715 FAE-----EFTQNKQSDVLEWGE-VVKELSVVEQIKRNYVDEDEDEE 759
 Db 892 FOESNDNTEFSTKYKVRNSLDEDESUKKELTKAEVVVK-----LDESESD 939

RESULT 14
 MYS2 DICDI
 ID_MYS2 DICDI STANDARD; PRT; 2116 AA.
 AC P08759;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Myosin II heavy chain, non muscle.
 GN MHCA.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=3540939;
 RX MEDLINE=87092266; PubMed=3540939;
 RA Warrick H.M., de Lozanne A., Leibwand L.A., Spudich J.A.;
 RT "Conserved protein domains in a myosin heavy chain gene from
 Dictyostelium discoideum";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
 RN [2]
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
 RC STRAIN=AX2;
 RX MEDLINE=90353583; PubMed=2387408;
 RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippller J.,
 RA Gerisch G.;
 RT "Replacement of threonine residues by serine and alanine in a
 phosphorylatable heavy chain fragment of Dictyostelium myosin II";
 RL FEBS Lett. 269:239-243(1990).
 RN [3]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=88112226; PubMed=3828113;
 RA Wagle G., Noegel A., Scheel J., Gerisch G.;
 RT "Phosphorylation of threonine residues on cloned fragments of the
 Dictyostelium myosin heavy chain";
 RL FEBS Lett. 227:71-75(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345066; PubMed=7619795;
 RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutch K., Holden H.M.,
 RA Rayment I.;
 RT "X-ray structures of the myosin motor domain of Dictyostelium
 discoideum complexed with MgADP.BeFx and MgADP.AlF4";
 RL Biochemistry 34:8960-8972(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345067; PubMed=7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 truncated head of Dictyostelium discoideum myosin to 2.7-A
 resolution";
 RL Biochemistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=96206189; PubMed=8611530;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the

Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
 RL Biochemistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RX MEDLINE=97452590; PubMed=9305951;
 RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATGgammaS, and MgAMPPNP complexes
 of the Dictyostelium discoideum myosin motor domain";
 RL Biochemistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=98070605; PubMed=9405148;
 RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg.2(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 Dictyostelium discoideum myosin motor domain";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14628; AAA33227.1; ..
 CC PIR; A26655; A26655.
 CC PDB; 1MNA; 03-DEC-97.
 CC PDB; 1MMD; 17-AUG-96.
 CC PDB; 1MMG; 03-DEC-97.
 CC PDB; 1MMN; 03-DEC-97.
 CC PDB; 1MND; 17-AUG-96.
 CC PDB; 1MNE; 23-AUG-96.
 CC PDB; 1VOM; 23-DEC-96.
 CC PDB; 1LVK; 28-JAN-98.
 CC PDB; 1DOY; 20-DEC-00.
 CC PDB; 1DOX; 20-DEC-00.
 CC PDB; 1DOZ; 20-DEC-00.
 CC PDB; 1DIA; 20-DEC-00.
 CC PDB; 1DIB; 20-DEC-00.
 CC PDB; 1DLC; 20-DEC-00.
 CC PDB; 1DMV; 20-DEC-00.
 CC PDB; 1PMW; 20-DEC-00.
 CC PDB; 1GBX; 17-JAN-01.
 CC PDB; 1JWY; 07-NOV-01.
 CC PDB; 1XK2; 07-NOV-01.
 CC DictyDB; DD01008; mhca.
 CC InterPro; IPR000048; IQ region.
 CC InterPro; IPR001609; myosin head.
 CC InterPro; IPR004009; Myosin_N.

Pfam; PF00612; IQ; 2.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761
FT DOMAIN 762 791
FT DOMAIN 817 2116
FT NP_BIND 179 186
FT DOMAIN 638 660
FT DOMAIN 738 752
FT MOD_RES 130 130
FT MOD_RES 678 678
FT MOD_RES 1823 1823
FT MOD_RES 1833 1833
FT MOD_RES 2029 2029
FT TURN 3 5
FT TURN 7 8
FT HELIX 10 15
FT HELIX 25 28
FT STRAND 34 37
FT TURN 41 42
FT STRAND 48 55
FT STRAND 59 63
FT STRAND 69 73
FT HELIX 74 76
FT STRAND 78 79
FT HELIX 83 85
FT TURN 86 87
FT STRAND 90 90
FT HELIX 91 93
FT HELIX 99 110
FT TURN 111 113
FT STRAND 116 119
FT TURN 120 121
FT STRAND 122 126
FT HELIX 137 142
FT TURN 143 145
FT HELIX 148 150
FT HELIX 155 169
FT TURN 170 170
FT STRAND 173 179
FT TURN 181 182
FT HELIX 185 200
FT HELIX 210 226
FT STRAND 227 228
FT TURN 234 235
FT STRAND 236 237
FT STRAND 240 247
FT TURN 249 250
FT STRAND 253 261
FT STRAND 265 268
FT TURN 269 269
FT TURN 273 274
FT STRAND 278 278
FT TURN 279 287
FT TURN 290 296
FT TURN 297 297
FT TURN 301 303
FT TURN 305 307
FT TURN 316 317
FT HELIX 320 334
FT TURN 335 335
FT HELIX 338 355
FT TURN 356 356
FT STRAND 360 360
FT STRAND 368 368
FT STRAND 373 382
FT HELIX 382

Query Match 5.1%; Score 201; DB 1; Length 2116;
Best Local Similarity 19.9%; Pred. No. 0.032;
Matches 165; Conservative 136; Mismatches 309; Indels 218; Gaps 33;

QY 14 LSLRVTAKELSLVNKNKSSAIVEIFESKY-----QAAAEETNMKKRSNTENLSQHFQKGT 68
DB 821 LKRRNFEKEI-----KEKEREILELKSNTLDSTQDKLEKSLKDTESNVLDLQRLQ-KAE 875
QY 69 LTVLKKKWNPGOLGAESHTSLRNSSTEIRHADHPPAEV-----TSHAASGA 116
DB 876 KETLKAMYDS-----KDALEAQKREIRVEDMESELDEKKLALENLQNKRSVEEK 927
QY 117 KADQEEQIHPRSLRSPPEALVQGRYPHIKQGEDLKHSTES-KWENCLGESRHEVEK- 174
DB 928 VRLDEELQEEQLKNTLEKLYVEELEEMKRYNDQSDTISRLEKIKDELQKEVEEL 987
QY 175 -SEISENTDASKIEKYNV---PLNRLKMFKEGPTQTILRAQ----- 215
DB 988 TESFSESKDKGVLEKTRVRLQSELDLDTVLDSBETKDSKLSLLROKKKLEELKOVQAL 1047
QY 216 -----SRASGRKISENSVSLDDLEIGPQQLSSSTFDSEKQESRRNLPLRSETSI 267
DB 1048 AAEATAKLAQEAANKL-QGEYT-----ELNEKFNSEVTARSNVEKSKTLESQLVAVNE 1102
QY 268 KDRMAKYOAAVSKSSSTN-----YTNELKASGGEEK-IHKMEQKENVPPGPEVCITHQE 321
DB 1103 LDEEKNRDALEKKKKALDAMLEEMKDQLESTGGKKSLYDLKVK-----QE 1149
QY 322 GEKISANENSLAVRSTPAEDDSRDSQVKSQVQVHPKPLSPDRASSLSSESSPKAMVK 381
DB 1150 SDMEALRNQISELQSTIAKLEKIKSTLEGEVAR-----LQGELEAEQLAKSNVEKQKK 1203
QY 382 FQAP-----ARETCVE--CQTVVPMERLLANQOVFHISCFRCSYCNKNKLSLGTVAS 431
DB 1204 VELDLDEKSAQLAEATAAKQALDKKKLQELSEVQT-----OLSEANNK----- 1249
QY 432 LHGRYICKPHFNQFLKSKGNVDEGHRPHKDLWASKNENEIEILERPAQLANARETSPSP 491
DB 1250 -----NVNSDS-----TNKHLSTFNVLKLEAEQKAKQALEKKR-- 1285
QY 492 QVEDAPIAKGVGLAASWEAK-ASSQQEKEDKPAETKRLRIAWPPPTTELGSSGSALEEGIK 550
DB 1286 -----LGESELKHVNEQLEEEKQKSNEKR-----KVDLEKEVSELKQOIE 1328
QY 551 MSKPKWPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTVAASFQSTSVKSPKTVSP 610
DB 1329 -----EEVASKKAVTE-----AKNKKSESELDEIKROYADVSVSRDKSVEQLKTLQ- 1373
QY 611 PIRKGSMSQESSESVGRVAERKQVENAKSKNG-----NVGKTTWQNK 656
DB 1374 --AKNEELRNTAEAEAG-----QLDRAERSKKKAEFDLEEA VKNLEETAKKVKAEKA 1424
QY 657 ESKGETGKRSKEGH-----SLEWENENLVENGA---DSDEDDNSFLKQSQSPQ 700
DB 1425 MKKAETDYRSTKSELDDAKNVSSSEQYVQIKLNEELSELSELSVLEAEADRCNSAIKAKTA 1484
QY 701 EPKSLNWSSFVD---NTFAEFTTQNKQSDVELWEGEVVVKLSVREQ 745
DB 1485 ESALSLKDEIDAANNAKAK---AERKSXELEVRVAEL---EESLEDK 1526

RESULT 15

MYH9_HUMAN

ID MYH9_HUMAN STANDARD; PRT; 1960 AA.

AC P35579; O60805;

DT 01-JUN-1994 (Rel. 29, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,

DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).

GN MYH9.

OS Homo sapiens (Human).

DR MIM; 605249; ..
DR GO; GO:0005960; C:non-muscle myosin; TAS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0036; IQ; 1.
DR Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family; Disease mutation;
KW Deafness.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 837 1926 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT VARIANT 93 93 N -> K (IN MHA).
FT VARIANT 702 702 /FTID-VAR 010791.
FT VARIANT 705 705 /FTID-VAR 010792.
FT VARIANT 1155 1155 R -> H (IN DNAA17).
FT VARIANT 1165 1165 /FTID-VAR 010794.
FT VARIANT 1424 1424 /FTID-VAR 010795.
FT VARIANT 1841 1841 D -> H (IN FTNS).
FT CONFLICT 53 55 E -> K (IN MHA).
FT CONFLICT 660 660 /FTID-VAR 010797.
FT CONFLICT 869 869 EAI -> RGH (IN REF. 3).
FT CONFLICT 931 931 T -> S (IN REF. 3).
FT CONFLICT 1240 1241 T -> M (IN REF. 4).
FT CONFLICT 1241 1241 C -> Y (IN REF. 4).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).

Query Match 5.1%; Score 200.5; DB 1; Length 1960;
Best Local Similarity 18.4%; Pred.No. 0.031;
Matches 167; Conservative 151; Mismatches 321; Indels 269; Gaps 34;

QY 14 LSLRVTA--ELSLVNNKSSAIVFISFYQAAEETNMEKRSNTENLSQ----- 62
DB 902 LAARLUTAKQEELEICHDLAEARVEEERCOHQAEE--KKMQQNIQEELEEEESAR 959
QY 63 ---HFRKGLTVLKKWENPGIAGSHSDLSLNSSTEIRHRADHPPEVTSAAAGAKAD 119
DB 960 QXLEKVTTEAKLKKLEBEQIILEDQNKLAKEKKLEDRI---AEFTNL-----TE 1010
QY 120 QEEIHPRLSPALVQGRYPHIKQEDIKDSTSKMKNCLGESRHEVEKSEISE 179
DB 1011 EEEKSLAKLKNKHEAMT-----TDLEERLRREKQORQELEKTRKLE----GD 1056
QY 180 NTDASGKIEKYNVPLNRLKW-MFEXGEPTQTKILRAQSRASG----PKISNSYSLDDL 234
DB 1057 STDLSQIAELQAQTAELKQAKLEEEELQALARVEEAAQKNMALKKIRLEESQISEL 1116
QY 235 --EIGFGQJSSSTFSEKNESRNLELPSLSETSIKDRMAKYQAQVSKSSSTNY---TN 289
DB 1117 QEDLESERASRNKAQKEDLGEELAEALKTELEDITLDTAAQQLRSKQBEVNLKXKL 1176
QY 290 ELKASGGEIKIHMKEQK-----NVPFGPEVCITHQEGE----- 323
DB 1177 EBEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRVKANLEKAKQT-LENERGELANEVKV 1235

QY 324 -----KISANENSLAVRSTPAEDSDSDSQVSEVQVPHPKPLSPDRAS-- 368
DB 1236 LLOQKGDSEHKRKVQAQLQELQVKNFE-----RVRTLADKVKTLQVLELONVTGLL 1289
QY 369 SLSESSPPKAMKTFQAPARETCVCOKTVYPMERLLANQQVHFHSCPRCSYCNKNKLSIGT 428
DB 1290 SOSDKSSKLTDFSA-----LESQLOQTQELQEE-----ROKLSUST 1329
QY 429 YASLHGRIYCKPHFNQLFKSKGNDE--GFGHRPHKDLWASKNENEIEILERPQALANARE 486
DB 1330 -----KLKQVEDEKNSFREOLEEBEAAKHLEKQIATLHAQVADMKK 1371
QY 487 TPHSPGVEDAPIAKVGVLAASMEAKASSQKEDKPAETKKLRIAMPPPTTELGGSSGSALE 546
DB 1372 K-----MEDS-----VGCLETAEEVKRKLQKLEGLSQRHEBKVAAYD----- 1409
QY 547 EGIKMSKPKWPEDEISKPEVPEDVDLDLKLRRSS-SLKERSRPF-----TVAASF 597
DB 1410 ---KLEKTKTLOQEL-----DOLLVDLDHQRSACNLEKKQKFFDQLLAEEXTISAKY 1460
QY 598 QSTSVKSPKTVSPPIRKWSMSEQSESVGGRVAE-----RKOVENAKSKKNGNVG 649
DB 1461 ABEEDRAEAAREKETKALSARALEEAMBOK-AELERLNKQFTEMEDLMSSK--DDVG 1517
QY 650 KTTW----- 653
DB 1518 KSVHELEKSKREALSQVVEEMKTLQEELEDELOATEDAKLRLEVNLMQAKQAFERDQGRD 1577
QY 554 -ONKESGKETGKRKSGHS-----LEMENENLVENGADSDDEDSNFL 694
DB 1578 EQSEKKKQLVRQVREMEAELEDERKORSNAVAARKKLEMDLKDLKLEAHIDSANKNRDEAI 1637
QY 695 KQSPQBPQKSLNWSFVQNTFA--EFTTQ-NQKSDQVELWEGEVV---KELSVBEQIKR 748
DB 1638 KQLRKLQAKQKDCMRLELDDTRASREELAQAKENEKKLKSMEAEMLQLEELAAERAKR 1697
QY 749 NRYVDEDE 756
DB 1698 QAQQRDE 1705

Search completed: January 6, 2004, 09:48:25
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:43:29 ; Search time 48 Seconds
(without alignments)
2509.864 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3927	100.0	759	21	AA197286 Lipid associated p
2	3927	100.0	759	22	ABUS2869 Human transcriptio
3	3927	100.0	759	22	AA197286 Human polypeptide
4	3925	99.9	759	22	AA197286 Human EPLIN (epith
5	3909.5	99.6	759	21	AA197286 Human ORFX-ORF2698
6	3083.5	78.5	600	22	AA197286 Human EPLIN (epith
7	2458	62.6	471	20	AA197286 Human breast sequ
8	2394	61.0	457	22	AA197286 Human breast tumou
9	2378	60.6	457	23	AA197286 Human Sterol regul

10	2011	51.2	476	22	ABG08559 Novel human diagno
11	1548	39.4	299	21	AA197286 Human secreted pro
12	1473	37.5	301	23	AB134206 Human cancer suppl
13	1254	31.9	243	21	AA197286 Human pancreatic c
14	1158	29.5	232	21	AA197286 Human secreted pro
15	1068	27.2	299	21	AA197286 Gene 41 human secr
16	889.5	22.7	235	20	AA197286 Human prostate tum
17	815	20.8	337	22	ABG24050 Novel human diagno
18	781	19.9	234	21	AA197286 Gene 5 human secre
19	583	14.8	152	22	AA197286 Human polypeptide
20	521.5	13.3	547	22	AA197286 Human secreted pro
21	521.5	13.3	547	23	AA197286 Human secreted pro
22	521.5	13.3	547	23	AA197286 Human 4141 protei
23	464.5	11.8	527	23	AB134206 Human breast tumou
24	464.5	11.8	527	23	AA197286 Human protein #5 c
25	393	10.0	88	21	AA197286 Human secreted pro
26	385.5	9.8	171	22	ABG04347 Novel human diagno
27	381	9.7	135	22	ABG12836 Haematopoietic ste
28	363	9.2	158	21	AA197286 Human inflammation
29	357	9.1	127	21	AA197286 Human NS protein s
30	357	9.1	173	23	ABG06053 Human NS protein s
31	357	9.1	173	23	ABG06054 Human NS protein s
32	357	9.1	173	23	ABG06055 Human NS protein s
33	357	9.1	173	23	ABG06056 Human NS protein s
34	357	9.1	173	23	ABG06057 Human NS protein s
35	357	9.1	173	23	ABG06058 Human NS protein s
36	354	9.0	111	22	ABG12835 Novel human diagno
37	325	8.3	217	22	AA197286 Human secreted pro
38	247	6.3	331	22	ABG08554 Novel human diagno
39	232.5	5.9	1793	22	AB134206 Drosophila melanog
40	231	5.9	134	21	AA197286 Pinus radiata tran
41	231	5.9	3111	22	ABG060327 Drosophila melanog
42	229	5.8	193	21	AA197286 Pinus radiata tran
43	226	5.8	993	22	AA197286 Human polypeptide
44	226	5.8	1312	22	AB134206 Human transcriptio
45	223	5.7	1312	21	AA197286 Retinoblastoma bin

ALIGNMENTS

RESULT 1

AA197286

ID AA197286 standard; Protein; 759 AA.

XX AA197286;

AC AA197286;

XX 03-JAN-2001 (first entry)

XX Lipid associated protein (LIPAP) 2766980CD1.

XX Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
KW antagonist; antibody; cardiovascular disease; neurological disease;
KW gastrointestinal disease; lipid metabolism; detection;
KW amplification; monitoring; hybridisation; antisense; triplex;
KW ribozyme; screening; immunoassay.

XX Homo sapiens.

XX Key Location/Qualifiers

FT	Modified-site	15	/note= "Potential phosphorylation site"
FT	Modified-site	19	/note= "Potential phosphorylation site"
FT	Modified-site	29	/note= "Potential phosphorylation site"
FT	Modified-site	48	/note= "Potential glycosylation site"
FT	Modified-site	55	/note= "Potential phosphorylation site"
FT	Modified-site	59	/note= "Potential phosphorylation site"
FT	Modified-site	68	/note= "Potential glycosylation site"

FT	Modified-site	658	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	662	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	686	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	698	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	706	
FT	Modified-site	/note= "Potential glycosylation site"	
FT	Modified-site	709	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	714	
FT	Modified-site	/note= "Potential phosphorylation site"	
FT	Modified-site	741	
FT	Modified-site	/note= "Potential phosphorylation site"	
XX			
PN	WO200049043-A2.		
XX			
PD	24-AUG-2000.		
XX			
PF	18-FEB-2000; 2000WO-US04160.		
XX			
PR	19-FEB-1999; 99US-0120703.		
PR	08-JUL-1999; 99US-0142762.		
XX			
PA	(INCY-) INCYTE PHARM INC.		
XX			
PI	Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;		
XX			
DR	WPI; 2000-549264/50.		
DR	N-PSDB; AAA53826.		
XX			
PT	New human lipid-associated proteins, nucleic acids, and antibodies,		
PT	useful for diagnosis, treatment and prevention of e.g. cardiovascular		
PT	disease		
XX			
PS	Claim 1; Page 76-77; 93pp; English.		
XX			
CC	Lipid-associated proteins (LIPAP) can be used for treating or		
CC	preventing disorders associated with decreased expression of LIPAP,		
CC	for screening for agonists or antagonists of LIPAP, and to raise		
CC	specific antibodies. Antagonists and antagonists of LIPAP are useful		
CC	for treating diseases associated with reduced or increased levels of		
CC	LIPAP, e.g. cardiovascular, neurological and gastrointestinal		
CC	diseases and disorders of lipid metabolism. Fragments of the nucleic		
CC	acid encoding LIPAP are useful for detection of full length coding		
CC	sequences, in hybridization and/or amplification assays or for		
CC	diagnosis or monitoring. Nucleotides encoding LIPAP are used		
CC	to screen for compounds that specifically modify LIPAP expression,		
CC	for recombinant production of LIPAP, in gene therapy, as a source of		
CC	therapeutic antisense, triplex-forming, or ribozyme agents and for		
CC	genomic mapping. Antibodies to the proteins are used for diagnosis		
CC	and monitoring of LIPAP-associated disease by immunoassay, as		
CC	antagonists, in competitive drug screens and for affinity		
CC	purification of natural LIPAP.		
XX			
SQ	Sequence 759 AA.		
SQ			

	Query Match	100.0%;	Score 3927;	DB 21;	Length 759;
	Best Local Similarity	100.0%;	Pred. No. 6.1e-299;		
	Matches 759;	Conservative	0;	Mismatches	0;
			Indels	0;	Gaps
					0;
QY	1	MESSPNNRRQWTSLSLRVTAKLSLVNKNKSSAIVEIFSKYOKAAEETNMEXKRNTENL	60		
Db	1	MESSPNNRRQWTSLSLRVTAKLSLVNKNKSSAIVEIFSKYOKAAEETNMEXKRNTENL	60		
QY	61	SOHFRKGTLTVLKKKWNPGLSAESTSLRNSSTEIRHADHPHAEVTSHAASGAKAQ	120		
Db	61	SOHFRKGTLTVLKKKWNPGLSAESTSLRNSSTEIRHADHPHAEVTSHAASGAKAQ	120		
QY	121	EQCIHPRSLRSPBALVQGRVPHIKDGEDLKHSTESKQMCNCLGESRHEVEKSEISEN	180		

Db 121 BEQIHPRRLRSLPPALVQGRYPHIKDGEDLKDHSKMKMNCIGESRHEVEKSEISEN 180
Qy 181 TDASGKIEKYNVPLNRLKMKFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ 240
Db 181 TDASGKIEKYNVPLNRLKMKFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ 240
Qy 241 LSSSTFDSEKNESSRNRLPRLSETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKI 300
Db 241 LSSSTFDSEKNESSRNRLPRLSETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKI 300
Qy 301 HOMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Db 301 HOMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Qy 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Db 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Qy 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGNVDGFGHRRPHKDLWASKNENEELERPAQ 480
Db 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGNVDGFGHRRPHKDLWASKNENEELERPAQ 480
Qy 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIANPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIANPPPTLGS 540
Qy 541 SGSALEEGIKMSKPKWPPDEISKEPEVEDVDLDLKLRRSSSLKERSRPTVAASTQST 600
Db 541 SGSALEEGIKMSKPKWPPDEISKEPEVEDVDLDLKLRRSSSLKERSRPTVAASTQST 600
Qy 601 SVKSPKTVSPPIRKGWSMESEESVGGRAERKOVENAKASKNGNKGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKGWSMESEESVGGRAERKOVENAKASKNGNKGKTTWQNKESKG 660
Qy 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQBPKSLNWSSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQBPKSLNWSSFVDNTFAEFT 720
Qy 721 TQNKSQDVELWEGEVVKLSVEEQIKNRYVDEDEDEE 759
Db 721 TQNKSQDVELWEGEVVKLSVEEQIKNRYVDEDEDEE 759

RESULT 2
ABUS2869
ID ABUS2869 standard; Protein; 759 AA.
XX AC ABUS2869;
XX DT 14-APR-2003 (first entry)
XX DE Human transcription factor from DKFzphutcl_18c19.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB01496.
XX PR 18-AUG-1999; 99US-0149499.
XX PR 28-SEP-1999; 99US-0156503.
XX PA (GBHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI: 2001-327840/34.
XX DR N-PSDB; ABX71307.

PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX Claim 21; Page 450-451; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a
CC polypeptide described in the disclosure of the invention.
XX

Seq Sequence 759 AA;

Query Match 100.0%; Score 3927; DB 22; Length 759;
Best Local Similarity 100.0%; Pred. No. 6.1e-299;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESSPFNRRRMTSLSRVTAKELSLVNKNKSSAIVEIFSKYKAAABETNMEKRSNTENL 60
Db 1 MESSPFNRRRMTSLSRVTAKELSLVNKNKSSAIVEIFSKYKAAABETNMEKRSNTENL 60
Qy 61 SQHFRKCTLTVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPPAEVTSHAASGAKAQ 120
Db 61 SQHFRKCTLTVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPPAEVTSHAASGAKAQ 120
Qy 121 EQIHPRSLRSPPEALVQGRYPHIKDGEDLKDHSKMKMNCIGESRHEVEKSEISEN 180
Db 121 EQIHPRSLRSPPEALVQGRYPHIKDGEDLKDHSKMKMNCIGESRHEVEKSEISEN 180
Qy 181 TDASGKIEKYNVPLNRLKMKFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ 240
Db 181 TDASGKIEKYNVPLNRLKMKFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ 240
Qy 241 LSSSTFDSEKNESSRNRLPRLSETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKI 300
Db 241 LSSSTFDSEKNESSRNRLPRLSETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKI 300
Qy 301 HKMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Db 301 HKMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Qy 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Db 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Qy 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGNVDGFGHRRPHKDLWASKNENEELERPAQ 480
Db 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGNVDGFGHRRPHKDLWASKNENEELERPAQ 480
Qy 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIANPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIANPPPTLGS 540
Qy 541 SGSALEEGIKMSKPKWPPDEISKEPEVEDVDLDLKLRRSSSLKERSRPTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPDEISKEPEVEDVDLDLKLRRSSSLKERSRPTVAASFQST 600
Qy 601 SVKSPKTVSPPIRKGWSMESEESVGGRAERKOVENAKASKNGNKGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKGWSMESEESVGGRAERKOVENAKASKNGNKGKTTWQNKESKG 660
Qy 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQBPKSLNWSSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQBPKSLNWSSFVDNTFAEFT 720
Qy 721 TQNKSQDVELWEGEVVKLSVEEQIKNRYVDEDEDEE 759
Db 721 TQNKSQDVELWEGEVVKLSVEEQIKNRYVDEDEDEE 759

RESULT 3

AAAM39013
ID AAAM39013 standard; Protein; 759 AA.

XX

AC AAAM39013;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 2159.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

XX

PR 25-APR-2000; 2000US-0552317.

XX

PR 09-JUL-2000; 2000US-0598042.

XX

PR 13-AUG-2000; 2000US-0620312.

XX

PR 03-AUG-2000; 2000US-0653450.

XX

PR 14-SEP-2000; 2000US-0662191.

XX

PR 19-OCT-2000; 2000US-0693036.

XX

PR 29-NOV-2000; 2000US-0727344.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX

PI Zhao QA, Zhou P, Goodrich R, Drnanac RT;

XX

DR WPI; 2001-442253/47.

XX

DR N-PSDB; AAI58169.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders

XX

PT such as central nervous system injuries -

XX

PS Example 4; SEQ ID NO 2159; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX

CC the encoded polypeptides (AAAM39013-AAAM42213) with nootropic,

XX

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX

CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX

CC of the invention may be used to treat diseases of the peripheral nervous

XX

CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX

CC localised neuropathies and central nervous system diseases, such as

XX

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX

CC utilisation of the activities such as: Immune system suppression,

XX

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

QY 1 MESSPFNRRQWTSLSLRTAKELSLVNNKSSAIVEIFSKYQKAAAEETNNKKRSNTENL 60
DB 1 MESSPFNRRQWTSLSLRTAKELSLVNNKSSAIVEIFSKYQKAAAEETNNKKRSNTENL 60
QY 61 SOHERKGTLTVLKKWENPGLGAESHTDSLNRNSTEIRHADHPPAEVTSHAASGAKAQD 120
DB 61 SOHERKGTLTVLKKWENPGLGAESHTDSLNRNSTEIRHADHPPAEVTSHAASGAKAQD 120
QY 121 EQIHRPSRLRSPPPEALVQGRYPHIKDGEDLKHSTESKMNENCLGESRHEVEKSEISEN 180
DB 121 EQIHRPSRLRSPPPEALVQGRYPHIKDGEDLKHSTESKMNENCLGESRHEVEKSEISEN 180
QY 181 TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
DB 181 TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSKNESSRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGBIKI 300
DB 241 LSSSTFDSKNESSRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGBIKI 300
QY 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDSQVSKSEVQPVHPKP 360
DB 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDSQVSKSEVQPVHPKP 360
QY 361 LSPDSRASSLSSESPPKAMKKFOAPARETCVECKTYVPMERLLANQOVFHSICFRCSYC 420
DB 361 LSPDSRASSLSSESPPKAMKKFOAPARETCVECKTYVPMERLLANQOVFHSICFRCSYC 420
QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKNGYDEGFGHRPHKOLWASKNENEELERPAQ 480
DB 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKNGYDEGFGHRPHKOLWASKNENEELERPAQ 480
QY 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASQSQEKEDKPAETKKLR.AWPPPTLGS 540
DB 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASQSQEKEDKPAETKKLR.AWPPPTLGS 540
QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFQST 600
DB 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPFPTVAASFQST 600
QY 601 SVKSPKTVSPPIRKGMSESESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKG 660
DB 601 SVKSPKTVSPPIRKGMSESESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVNTFAFEFT 720
DB 661 ETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVNTFAFEFT 720
QY 721 TQNXQSDQVLEWGEVVKELSVVEQIKENRYDEDEDEE 759
DB 721 TQNXQSDQVLEWGEVVKELSVVEQIKENRYDEDEDEE 759
RESULT 4
AAB67701
ID AAB67701 standard; Protein; 759 AA.
XX AAB67701;
AC AAB67701;
XX 11-JUN-2001 (first entry)
DT Human EPLIN (epithelial protein lost in neoplasm)-beta isoform.
XX Human EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
XX gene therapy; cancer.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 278
FT Misc-difference /note= "Val encoded by GCG"
FT Misc-difference 344

FT /note= "Arg encoded by CCAGGT"
 FT Misc-difference 495
 FT /note= "Asp encoded by AAT"
 FT Misc-difference 511
 FT /note= "Lys encoded by AGG"
 FT Misc-difference 622
 FT /note= "Ser encoded by AAT"
 FT Misc-difference 658
 FT /note= "Ser encoded by TTT"
 FT Misc-difference 679
 FT /note= "Ile encoded by TTT"
 FT Misc-difference 698
 FT /note= "Ser encoded by TTT"
 FT Misc-difference 704
 FT /note= "Ser encoded by TTT"
 XX
 XX WO200118019-A1.
 XX
 XX 15-MAR-2001.
 XX
 XX 08-SEP-2000; 2000WO-US24689.
 XX
 XX 08-SEP-1999; 99US-0153024.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Chang DD, Maul RS;
 XX
 XX WPI; 2001-244555/25.
 XX N-PSDB; AAF55697.
 XX
 XX New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
 XX prognostic and therapeutic applications over the course of cell
 XX proliferative disorders associated with EPLIN -
 XX
 XX Claim 1; Page 44; 59pp; English.
 XX
 XX The present sequence represents a human EPLIN (epithelial protein lost in
 XX neoplasm)-beta isoform. The specification also describes EPLIN-alpha.
 XX EPLIN is a tumour suppressor protein, whose expression is altered in
 XX multiple common human tumour types. EPLIN nucleic acids and proteins are
 XX used in screening assays to detect molecules that specifically bind to
 XX EPLIN nucleic acids, proteins or derivatives and thus have potential use
 XX as agonist or antagonist of EPLIN, in particular molecules that affect
 XX cell proliferation. Thus the assays are useful for screening molecules
 XX with potential utility as anticancer drugs or lead compounds for drug
 XX development. EPLIN nucleic acids, proteins are useful for detecting a
 XX cell proliferative disorder in a subject. EPLIN polynucleotides are
 XX useful in gene therapy techniques. EPLIN is useful as a marker that
 XX can be diagnostically, prognostically and therapeutically used over
 XX the course of a cell proliferative disorder associated with EPLIN.
 XX
 XX Sequence 759 AA;
 XX
 XX Query Match 99.9%; Score 3925; DB 22; Length 759;
 XX Best Local Similarity 99.9%; Pred. No. 8.8e-299;
 XX Matches 758; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 1 MESSPPNRRQWTSLSLRVAKELSLVKNKSSAIVLFSKYKAAETNNKKRSNTENL 60
 XX
 XX 1 MESSPPNRRQWTSLSLRVAKELSLVKNKSSAIVLFSKYKAAETNNKKRSNTENL 60
 XX
 XX 61 SQHFRKGTTLVKKKWNENPGIAGESHTDSLRSNSTEIRHRADHPPEVTSAAASGAKADQ 120
 XX
 XX 61 SQHFRKGTTLVKKKWNENPGIAGESHTDSLRSNSTEIRHRADHPPEVTSAAASGAKADQ 120
 XX
 XX 121 EQHPRSRRLRSPPEALVQGRYPHIKQEDLKHSTESKKWENCLGESRHEVEKSEISEN 180
 XX
 XX 121 EQHPRSRRLRSPPEALVQGRYPHIKQEDLKHSTESKKWENCLGESRHEVEKSEISEN 180
 XX
 XX 181 TDASGKTEKYNVPLNRLKWMFEKCEPTQTKILRAQSRSSASGRKISENSYSLDLEIGPGQ 240
 XX
 XX 181 TDASGKTEKYNVPLNRLKWMFEKCEPTQTKILRAQSRSSASGRKISENSYSLDLEIGPGQ 240

QY 241 LSSSTFDESEKNSRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
 Db 241 LSSSTFDESEKNSRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
 QY 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDQSVKSEVQPVHPKP 360
 Db 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDQSVKSEVQPVHPKP 360
 QY 361 LSPDRASSLSSESPPKAMKKFQAPARETCVEQKTVYPMERLLANQQVPHISCFRCYSY 420
 Db 361 LSPDRASSLSSESPPKAMKKFQAPARETCVEQKTVYPMERLLANQQVPHISCFRCYSY 420
 QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNVDEGFGHPRPHKDLWASKNENEBILRPAQ 480
 Db 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNVDEGFGHPRPHKDLWASKNENEBILRPAQ 480
 QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRITAWPPPTLGS 540
 Db 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRITAWPPPTLGS 540
 QY 541 SGSALBEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFVVAASFQST 600
 Db 541 SGSALBEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFVVAASFQST 600
 QY 601 SVKSPKTVSPPIRKGWSMGESESYGGRVAERKQVENAKASKKNGNVGKTTWQNKESKG 660
 Db 601 SVKSPKTVSPPIRKGWSMGESESYGGRVAERKQVENAKASKKNGNVGKTTWQNKESKG 660
 QY 661 ETGKRSKEGHSLEMEENENENENGADSDDEDNSFLKQSPQEPKSLNWSFVDMNTFAEFT 720
 Db 661 ETGKRSKEGHSLEMEENENENENGADSDDEDNSFLKQSPQEPKSLNWSFVDMNTFAEFT 720
 QY 721 TQNKQSQDVELWEGEVVVEEQIKRNRYYDEDEDEE 759
 Db 721 TQNKQSQDVELWEGEVVVEEQIKRNRYYDEDEDEE 759
 XX
 XX RESULT 5
 XX AAB42934
 XX ID AAB42934 standard; Protein; 760 AA.
 XX
 XX AC AAB42934;
 XX
 XX DT 08-FEB-2001 (first entry)
 XX
 XX DE Human ORFX ORF2698 polypeptide sequence SEQ ID NO:5396.
 XX
 XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX KW vulnery; antipsoriatic; antiparkinsonian; neurotrophic; neuroprotective;
 XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 XX KW antiviral; antibacterial; antifungal; antineoplastic; antitumor;
 XX KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
 XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX KW thrombosis; contraceptive.
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO200058473-A2.
 XX
 XX PD 05-OCT-2000.
 XX
 XX PF 31-MAR-2000; 2000WO-US08621.
 XX
 XX PR 31-MAR-1999; 99US-0127607.
 XX 02-APR-1999; 99US-0127636.

PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN -
XX
PS Claim 1; Page 43; 59pp; English.

The present sequence represents a human EPLIN (epithelial protein lost in neoplasm)-alpha isoform. The specification also describes EPLIN-beta. EPLIN is a tumour suppressor protein, whose expression is altered in multiple common human tumour types. EPLIN nucleic acids and proteins are used in screening assays to detect molecules that specifically bind to EPLIN nucleic acids, proteins or derivatives and thus have potential use as agonist or antagonist of EPLIN, in particular molecules that affect cell proliferation. Thus the assays are useful for screening molecules with potential utility as anticancer drugs or lead compounds for drug development. EPLIN nucleic acids, proteins are useful for detecting a cell proliferative disorder in a subject. EPLIN polynucleotides are useful in gene therapy techniques. EPLIN is useful as a marker that can be diagnostically, prognostically and therapeutically used over the course of a cell proliferative disorder associated with EPLIN.

Sequence 600 AA;

Query Match	78.5%;	Score 3083.5;	DB 22;	Length 600;
Best Local Similarity	99.5%;	Pred. No. 6.4e-233;		
Matches 597; Conservative	1;	Mismatches 1;	Indels 1;	

QY		161	MENCJGESRHEVEKSEI	S	E	N	T	D	A	S	G	K	I	E	K	Y	N	V	P	L	N	R	L	K	O	M	F	E	K	G	E	P	T	K	I	L	R	A	Q	S	R	S	A	S	220
Db		1	MENCJGESRHEVEKSEI	S	E	N	T	D	A	S	G	K	I	E	K	Y	N	V	P	L	N	R	L	K	O	M	F	E	K	G	E	P	T	K	I	L	R	A	Q	S	R	S	A	S	60

Qy 221 GRKISSENSYLDLEIGFGQLSSTFDSEKNESRRNLEIPLRSETSIKORMAKYQAASVK 280
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 GRKISSENSYLDLEIGFGQLSSTFDSEKNESRRNLEIPLRSETSIKORMAKYQAASVK 120

[illegible][illegible]

QY 400 MERLLANQQVPHISCFRCSYCNKKLSLGTYSALHGRIRYCKPHFNQLFKSKGNYDEGFGHR 459
 |||||
 241 MERLLANQQVPHISCFRCSYCNKKLSLGTYSALHGRIRYCKPHFNQLFKSKGNYDEGFGHR 300
 Db

[illegible]

Qy 520 DKPAETKKLR¹AWPPPT²ELGSSGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLR 579
 |||||
 361 DKPAETKKLR¹AWPPPT²ELGSSGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLR 420

Qy	580	RSSSLKERSRPPTVAASFQSTSVKSPKTVSPPIRKGWSNSEQSESVGGRVAERKQVENA	639
	421	RSSSLKERSRPPTVAASFQSTSVKSPKTVSPPIRKGWSNSEQSESVGGRVAERKQVENA	480
Db			

QY 640 KASKKGNVGKTTWQNKESGKGTGRSKEGHSLEMEENINVENGCADSDDDNSFLKQSP 699

Db 481 KASKKGNVGKTTWQNKESGKGTGRSKEGHSLEMEENINVENGCADSDDDNSFLKQSP 540

QY 700 QEPKSLNWSSFDVNTFAEEFTTQKSDQDVELWEGEVVKLSVEEQIKRNRYYDEDEDE 759
Db 541 QEPKSLNWSSFDVNTFAEEFTTQKSDQDVELWEGEVVKLSVEEQIKRNRYYDEDEDE 600

RESULT 7
AAAY48487
ID AAAY48487 standard; Protein: 471 AA.

AA
AC
AAAY48487;

XX	
DT	08-DEC-1999 (first entry)
DE	
DE	Human breast tumour-associated protein 32.
DE	
KW	Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW	medicaments; gene therapy; treatment; fat metabolism.
XX	
OS	Homo sapiens.
PN	DE19813835-A1.
PP	23-SEP-1999.
PD	
XX	
XX	20-MAR-1998; 98DE-1013835.
XX	
PR	20-MAR-1998; 98DE-1013835.
XX	
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.
PI	
FI	Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
DR	WPI; 1999-528979/45.
DR	N-PSDB; AAZ33566.
PT	
FT	Human nucleic acid sequences and protein products from normal breast
PFT	tissue, useful for breast cancer therapy -
XX	
PS	Claim 28; 173; 206pp; German.
CC	
CC	This invention describes novel human nucleic acid sequences from normal
CC	breast tissue which have cytostatic activity. The nucleic acid sequences
CC	can be used to produce and isolate full-length gene sequences. They can
CC	be used to express proteins, which can be used as tools to find an
CC	activity against breast cancer. The sequences can be used in sense or
CC	antisense form. They are especially useful for medicaments for gene
CC	therapy to treat breast cancer and for treating illnesses associated
CC	with fat metabolism. AA#48456-Y48539 represent protein fragments encoded
CC	by the expressed sequence tags described in the method of the invention.
XX	
SQ	Sequence 471 AA;

Query Match 62.6%; Score 2458; DB 20; Length 471;
Best Local Similarity 99.8%; Pred. No. 5.3e-184;
Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0

Qy 289 NELKASGGEIKIHMEQKENVPPGPVCITHQEGETISANENSLAVRSTPAEDDSRDQV 348
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 1 NELKASGGEIKIHMEQKENVPPGPVCITHQEGETISANENSLAVRSTPAEDDSRDQV 60

	Qy	349 KSEVQP VHPKPLSPDRASSLSESSPPKAMKKFQAPARETCVECQTVPMERLLANQQ 408
	Dd	61 KSEVQP VHPKPLSPDRASSLSESSPPKAMKKFQAPARETCVECQTVPMERLLANQQ 120

[illegible]

469 NENEILERPAQLANARETHPSPGVEDAPIAKGGV^{547K}LAA^{548K}SM^{549K}EAKASSQOEKEDPAETKKL 528
181 NENEILERPAQLANARETHPSPGVEDAPIAKGGV^{547K}LAA^{548K}SM^{549K}EAKASSQOEKEDPAETKKL 240

[illegible]

	OY	589 RPFTVAASFQSTSVKSPKTVPPIRKGWSMSEQEESVSGRVAERKOVENAKASKONGNV 648
	bD	301 RPFTVAASFQSTSVKSPKTVPPIRKGWSMSEQEESVSGRVAERKOVENAKASKONGNV 360

649	GKTTWQNKSKGETCKRSKEGHSLEWENENLVENGADSDDEDNSFLKQSQPOEPKSLNWS	708
361	GKTTWQNKSKGETCKRSKEGHSLEWENENLVENGADSDDEDNSFLKQSQPOEPKSLNWS	420

20

QY 709 SFVDNTFAEFTTQKSDQVELWEGVVKLSVEEQIKRNRYDEDEDE 759
AAB95477
DB 421 SFVDNTFAEFTTQKSDQVELWEGVVKLSVEEQIKRNRYDEDEDE 471

RESULT 8
AAB95477
ID AAB95477 standard; Protein; 457 AA.

XX AAB95477;

AC 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:17992.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 17992; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides,

Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 303 MBOKENVPPGVEVCITTHOEGEKISANENSLAVRSTPAEDDSRDQSKSEVOQPVHPKPLS 362
DB 1 MEOKENVPPGVEVCITTHOEGEKISANENSLAVRSTPAEDDSRDQSKSEVOQPVHPKPLS 60
QY 363 PDSRASSLSSESPPKAMKKFOAPARETCVEQKTVYPMERLLANQOVFHHISCFRCSCYNN 422
DB 61 PDSRASSLSSESPPKAMKKFOAPARETCVEQKTVYPMERLLANQOVFHHISCFRCSCYNN 120
QY 423 KLSLGTYSALHGRYICKPHFNQLFKSKGNDEGEFGHPRHKLWASKNENBEILERPQOLA 482
DB 121 KLSLGTYSALHGRYICKPHFNQLFKSKGNDEGEFGHPRHKLWASKNENBEILERPQOLA 180
QY 483 NARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSG 542
DB 181 NARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSG 240
QY 543 SALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPFTVAASFQSTSV 602
DB 241 SALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPFTVAASFQSTSV 300
QY 603 KSPKTVSPPIRKGMSSMSQSEESVGRVAERKQVENAKASKKNGNVGKTTWONKESKGET 662
DB 301 KSPKTVSPPIRKGMSSMSQSEESVGRVAERKQVENAKASKKNGNVGKTTWONKESKGET 360
QY 663 GKRSKEGHSLEMNENLVENGADSDSDNSFLKQOSPQEPKSLWSSFVDNTFAEFTTQ 722
DB 361 GKRSKEGHSLEMNENLVENGADSDSDNSFLKQOSPQEPKSLWSSFVDNTFAEFTTQ 420
QY 723 NQKSDQVELWEGVVKLSVEEQIKRNRYDEDEDE 759
DB 421 NQKSDQVELWEGVVKLSVEEQIKRNRYDEDEDE 457

RESULT 9

AAU10979

ID AAU10979 standard; Protein; 457 AA.

XX AAU10979;

AC AAU10979;

XX 12-MAR-2002 (first entry)

DT Human sterol regulatory element binding protein 3.

DE Human; sterol regulatory element binding protein 3; hSRBEP-3;

KW hypothalamus.

XX Homo sapiens.

OS CN1309182-A.

PN 22-AUG-2001.

PD 17-FEB-2000; 2000CN-0111698.

XX 17-FEB-2000; 2000CN-0111698.

XX (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.

PA Li Y, Xu S, Ren S;

XX WPI; 2002-011822/02.

DR N-PSDB; AAS18588.

XX Cholesterol regulatory factor binding protein and its coding sequence -

PT Claim 2; Page 14; 27pp; Chinese.

PS The invention relates to a novel human sterol regulatory element binding

XX protein 3 (hSRBEP-3) expressed in human normal hypothalamic tissue and

CC its coding sequence. Also described is the process for preparing the

CC protein and nucleic acid sequence, and the method for detecting hSRBEP-3

CC nucleic acid sequence and polypeptides. The present sequence represents
CC the amino acid sequence of human HSREBP-3 as described in the invention.
XX
SQ Sequence 457 AA;

Query Match 60.6%; Score 2378; DB 23; Length 457;
Best Local Similarity 99.3%; Pred. No. 9, 5e-178;
Matches 454; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 303 MQQENVPVPGVCITHQGEKISANENSLAVRSTPAEDDSRDQVSEVQVHPKPLS 362
Db 1 MQQENVPVPGVCITHQGEKISANENSLAVRSTPAEDDSRDQVSEVQVHPKPLS 60
QY 363 PDSRASSLSSESPKAMKFKQAPARETCVECKTVPYMERLLANQQVPHISCFRCSCYNN 422
Db 61 PDSRASSLSSESPKAMKFKQAPARETCVECKTVPYMERLLANQQVPHISCFRCSCYNN 120
QY 423 KLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENELRPAQLA 482
Db 121 KLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENELRPAQLA 180
QY 483 NARETPHSPGVDEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSG 542
Db 181 NARETPHSPGVDEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSG 240
QY 543 SALEEGIKMSKPKWPPDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQSTSV 602
Db 241 SALEEGIKMSKPKWPPDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQSTSV 300
QY 603 KSPKTVSPPIRKGWSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESGET 662
Db 301 KSPKTVSPPIRKGWSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESGET 360
QY 663 GKRSEKHSLEMENLVENGADSDSDNSFLKQOSPQEPKSLNWSSFDNTFAEFTTQ 722
Db 361 GKRSEKHSLEMENLVENGADSDSDNSFLKQOSPQEPKSLNWSSFDNTFAEFTTQ 420
QY 723 NQSQDVELWEVVKLSVEEQIKRNYDEDEDEE 759
Db 421 NQSQDVELWEVVKLSVEEQIKRNYDEDEDEE 457

RESULT 10
ABG08559
ID ABG08559 standard; Protein: 476 AA.
XX AC ABG08559;
XX DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #8550.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS72746.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID NO 38918; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 476 AA;
SQ
Query Match 51.2%; Score 2011; DB 22; Length 476;
Best Local Similarity 90.4%; Pred. No. 6, 3e-149;
Matches 405; Conservative 7; Mismatches 28; Indels 8; Gaps 7;
QY 320 QEKEKISANENSLAVRSTPAEDDS-RDSQVKSEVQVHPKPLSPDSRASSLSSESPPKA 378
Db 29 QEKEKISANENSLAVRSTPAEDDSPGDSQVKSEVQVHPKPLSPDSRASSLSSESPPKA 88
QY 379 MKXFPAPARETCVECKTVPYMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYIC 438
Db 89 MKXFPAPARETCVECKTVPYMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYIC 148
QY 439 KPHFNQLFKSKGNYDEGFGHRPHKDL-WASKNENELRPAQLANARETPHSPGVDEAP 497
Db 149 KPHFNQLFKSKGNYDEGFGHRPHKDLIGQAKMKTETILRQPOLANARETPHSPGVDEAP 208
QY 498 IAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWP 557
Db 209 IAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWP 268
QY 558 PEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRKGWS 617
Db 269 PEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRKGWS 328
QY 618 MSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKGETGK-RSKSEGHSELEME 676
Db 329 MSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKGETGKRRXKGVHSELEME 388
QY 677 ENLVENGADSD-EDNSF--LKQSQPQEPKSLNWS-SFVDN-TFAEEFTTQNKESQDVEL 731
Db 389 ENLVENGADSD-EDNSF--LKQSQPQEPKSLNWS-SFVDN-TFAEEFTTQNKESQDVEL 448
QY 732 WEQEVVKLSVEEQIKRNYDEDEDEE 759
Db 449 WEQEVVKLSVEEQIKRNYDEDEDEE 476

RESULT 11
AAB34206
ID AAB34206 standard; Protein: 299 AA.
XX
AC AAB34206;

XX DT 26-JAN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:174.
XX KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
XX KW antineumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
XX KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;
XX KW infection; ocular disorder; wound healing; skin aging; food additive;
XX KW preservative.
XX OS Homo sapiens.
XX PN WO200056755-A1.
XX PD 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06830.
XX PR 19-MAR-1999; 99US-0125361.
XX PR 10-DEC-1999; 99US-0169910.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX DR WPI; 2000-587661/55.
XX Nw isolated nucleic acid molecules encoding 49 human secreted proteins
XX used for preventing, treating or ameliorating medical conditions, for
XX diagnosing pathological conditions or as food additives or
XX preservatives -
XX Disclosure; Page 412-413; 419pp; English.
XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX AAB34216 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissue
XX and cells the genes are expressed in. Examples of activities include:
XX antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX and vulnerary. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
XX cancer of the breast or liver, cardiovascular disorders, cerebrovascular
XX disorders, angiogenesis, nervous system disorders, infections caused by
XX bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX also be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
XX in the exemplification of the present invention.

QY 514 SQBQEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDL 573
DB 61 SQBQEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDL 120
QY 574 DLKLRSSSLKERSRPFPTVAASFQSTSVKSPKTVSPPIRKGMSMSQSESVCGRAVER 633
DB 121 DLKLRSSSLKERSRPFPTVAASFQSTSVKSPKTVSPPIRKGMSMSQSESVCGRAVER 180
QY 634 KOVENAKASKNGNVGKTTWQNKESKGETGKRSKEGHSLEMNENLVENGADSDSDNSF 693
DB 191 KOVENAKASKNGNVGKTTWQNKESKGETGKRSKEGHSLEMNENLVENGADSDSDNSF 240
QY 694 LKQSPQEPKSLNWSSFVDNTFAEEFTTQNKSOODVELWEGEVVVKLSVEEQIKRNRYY 752
DB 241 LKQSPQEPKSLNWSSFVDNTFAEEFTTQNKSOODVELWEGEVVVKLSVEEQIKRNRYY 299
RESULT 12
AAB56420
ID ABB56420 standard; Protein; 301 AA.
XX AC ABB56420;
XX DT 21-FEB-2002 (first entry)
XX DE Human cancer suppressor protein PP624.
XX KW Human; cancer suppressor; disease; cancer.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 21 /label= unknown
FT Misc-difference 40 /note= "Encoded by RCA"
FT Misc-difference 40 /label= unknown
FT Misc-difference 43 /note= "Encoded by TRC"
FT Misc-difference 43 /label= unknown
FT Misc-difference 43 /note= "Encoded by AKC"
CN1313297-A.
PD 19-SEP-2001.
PF 09-MAR-2000; 2000CN-0111948.
PR 09-MAR-2000; 2000CN-0111948.
PA (SHAN-) SHANGHAI INST ONCOLOGY.
XX Gu J, Yang S;
XX WPI; 2002-042185/06.
XX N-PSDB; ABI98973.
XX Human protein able to suppress growth of cancer cells and its coding
XX sequence -
XX Claim 1; Page 20 Disclosure; 37pp; Chinese.
XX The invention relates to novel human proteins (ABB56417-ABB56425) with
XX cancer suppressing function, the encoding polynucleotides
XX (ABI98970-ABI98978), the process for preparing the polypeptide, the
XX application of the polypeptide in treating diseases such as cancer, the
XX antagonist of the polypeptide and its medical function and the
XX application of the polynucleotide.
SQ Sequence 301 AA;

PA (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI: 2000-619227/59.
DR N-PSDB; AAC93483.
XX New nucleic acid molecules encoding 49 human secreted proteins for
PT diagnosing, preventing or ameliorating medical conditions and used for
PT food additives or preservatives -
XX
XX Disclosure; Page 22; 516pp; English.
XX
XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
XX human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
XX AAB51927 represent alternative polypeptides encoded by the genes, and
XX amino acid sequences with which they share homology. The genes and
XX proteins have activities dependent on the tissues and cells in which they
XX are expressed. Examples of their activities include immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotrophic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
XX virucide; fungicide; ophthalmological; and vulnerary. The secreted
XX proteins, polynucleotides, antagonists and agonists may be useful in
XX treating, preventing and/or diagnosing diseases and disorders such as
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiodysplasia, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,
XX carbohydrate, vitamins, minerals, cofactors and other nutritional
XX components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
XX used in the isolation and characterisation of the proteins and
XX polynucleotides of the invention.
XX
SQ Sequence 232 AA;
Query Match 29.5%; Score 1158; DB 21; Length 232;
Best Local Similarity 100.0%; Pred. No. 1.7e-82;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 473 EILERRAQLANARETHSPGVEDAPITAKGVLAASMEAKASSQOEKEDPAETTKLRIAM 532
Db 1 EILERRAQLANARETHSPGVEDAPITAKGVLAASMEAKASSQOEKEDPAETTKLRIAM 60
QY 533 PPTELSSGSSALEEGIKMSKPKWPPDETSKPEVPEDVDLDDKLLRSSLSLKERSRPPT 592
Db 61 PPTELSSGSSALEEGIKMSKPKWPPDETSKPEVPEDVDLDDKLLRSSLSLKERSRPPT 120
QY 593 VAASFQSTSVKSPKTSVPPPIRKQWMSSESESVGGRVAERKQVENAKASKXNGNVGKTT 652
Db 121 VAASFQSTSVKSPKTSVPPPIRKQWMSSESESVGGRVAERKQVENAKASKXNGNVGKTT 180
QY 653 WQNKESGEGTGRKSGKHSLEMEENLHVENGADSDDEDDNSFLKQSQP 699
Db 181 WQNKESGEGTGRKSGKHSLEMEENLHVENGADSDDEDDNSFLKQSQP 227

KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative.
XX Cricetulus griseus.
OS WO200056755-A1.
XX
XX 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06830.
XX
XX 19-MAR-1999; 99US-0125361.
XX 10-DEC-1999; 99US-0169910.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI: 2000-587661/55.
XX
XX New isolated nucleic acid molecules encoding 49 human secreted proteins
XX used for preventing, treating or ameliorating medical conditions, for
XX diagnosing pathological conditions or as food additives or
XX preservatives -
XX
XX Disclosure; Page 411-412; 419pp; English.
XX
XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX AAB34216 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissue
XX and cells the genes are expressed in. Examples of activities include:
XX antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotrophic; cerebroprotective; neurotropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX and vulnerary. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
XX cancer of the breast or liver, cardiovascular disorders, cerebrovascular
XX disorders, angiogenesis, nervous system disorders, infections caused by
XX bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX also be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
XX in the exemplification of the present invention.
XX
SQ Sequence 299 AA;
Query Match 27.2%; Score 1068; DB 21; Length 299;
Best Local Similarity 71.0%; Pred. No. 3e-75;
Matches 213; Conservative 30; Mismatches 55; Indels 2; Gaps 2;
QY 454 EGFGRPHKDLWASKNENEILERRAQLANARETHSPGVEDAPITAKGVLAASMEAKAS 513
Db 1 EGFGRPHKDLWASKNENEILERRAQLANARETHSPGVEDAPITAKGVLAASMEAKAS 60
QY 514 SQQEKEDKPAETTKLRIAMPPPTTELSSGSSALEEGIKMSKPKWPPDETSKPEVPEDVDL 573
Db 61 PQREKEDKPAETTKLRIAMPPPPAEELSSGSSGTMEEGFKSGKPKPPPEDETSKTAALGVDL 120

QY 574 DLAKLRSSSLKERSRPFTVAASFQSTSVKSPKTVSPPIRKGWSMSEQSEESVGGPVAER 633
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 DLAKLRSSSLKERSRPFTVAASFRTSSLUKSPKTLSPIRKGWSMSEQNEESGGIVTAG 179
QY 634 KOVENAKASKKNGNUGKTTWQNKES-KGETGKRSKEGHSLENNENLVENGADSDDEDDNS 692
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
180 KOTENAKAPCKKENVRKSRWOREDSAREEVARRSKDIHGVELESENFIENGANIYEGDSC 239
QY 693 FLKQSPQBPKSLNWSFVDNTFAEFTTONQKSQDVELWEGEVVKELSVEEQIKENRY 752
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
240 DLVQSPPLDPKSPWSGFDVTPTTEFTTQNKSHDVGFRGEVVRVRELVEEQIKENRY 299

Search completed: January 6, 2004, 09:47:18
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 6, 2004, 13:51:06 ; Search time 13109 Seconds
(without alignments)
11562.298 Million cell updates/sec
Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcctcagagcagtaggtg.....atttgaaaaaaaaaaaaa 3705
Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0
Searched: 2888711 seqs, 2045481386 residues
Word size : 0
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_scs.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_ox.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sv.*
- 39: em_hgo_hum.*
- 40: em_hgo_mus.*
- 41: em_hgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3611	97.5	3664	6	AX086386
2	3611	97.5	3664	9	AX086386 Sequence
3	3548	95.8	3655	9	AF136911 Homo sapi
4	3114	84.0	3277	9	AF198454 Homo sapi
5	3098	83.6	3550	9	AX000335 Homo sapi
6	2647	71.4	2783	6	AF198455 Homo sapi
7	2647	71.4	2783	6	AX017487 Sequence
8	2386	64.4	2536	9	BD135177 Human nuc
9	2324	62.7	3348	9	BD135177 Homo sapi
10	2234	60.3	115345	9	AF157325 Homo sapi
11	1967	53.1	2121	9	AF157325 Homo sapi
12	1903	51.4	2667	9	BC008147 Homo sapi
13	1901	51.3	2164	9	BC008147 Homo sapi
14	1519	41.0	2207	6	AX000057 Homo sapi
15	1519	41.0	2207	6	BC001247 Homo sapi
16	1462	39.5	2793	9	BD160117 Primer fo
17	1296	35.0	1754	9	AX023649 Homo sapi
18	1109	29.9	28186	9	AK096172 Homo sapi
19	945	25.5	1021	6	AF218025 Homo sapi
20	714	19.3	777	11	AC139016 Homo sapi
21	604	16.3	641	6	AX012180 Sequence
22	390	10.5	390	6	G56637 SHGC-102035
23	343	9.3	343	6	AX011639 Sequence
24	343	9.3	343	6	BD113230 EST and e
25	325	8.8	389	6	AX378979 Sequence
26	276	7.4	279	6	AX379294 Sequence
27	248	6.7	413	6	BD025824 Sequence
28	238	6.4	250	11	AR269930 Sequence
29	236	6.4	532	6	AR270117 Sequence
30	212	5.7	247	6	GI1512 human STS S
31	209	5.6	231	6	BD154971 Primer fo
32	209	5.6	231	6	AX184947 Sequence
33	208	5.6	265	6	AX187815 Sequence
34	175	4.7	178	6	AX186750 Sequence
35	143	3.9	296	6	AX098162 Sequence
36	119	3.2	238	6	AX341839 Sequence
37	116	3.1	1109	9	BD027230 Sequence
38	87	2.3	213	6	AR269618 Sequence
39	86	2.3	86	6	AF245391 Homo sapi
40	42	1.1	840	6	AX198679 Sequence
41	36	1.0	425	10	AX209212 Sequence
42	31	0.8	2370	10	AX323411 Sequence
43	31	0.8	3232	10	AF307845
44	31	0.8	3997	10	BC031490
45	31	0.8	195301	2	AF307844 Mus muscu
					AC134548 Mus muscu

ALIGNMENTS

RESULT 1.
AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Wiemann,S.
TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 338 22-FEB-2001;
Pred. No. is the number of results predicted by chance to have a

German Human Genome Project (DE)
Location/Qualifiers
source 1.3664
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606" 911 t
BASE COUNT 1210 a 734 c 809 g
ORIGIN
Query Match 97.5%; Score 3611; DB 6; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 44 GCGTAGGTAGAGCGCGGACCTGTACAGGGCTGTAGCAGCGAGAGAAAGCGGC 103
Db 1 GCGTAGGTAGAGCGCGGACCTGTACAGGGCTGTAGCAGCGAGAGAAAGCGGC 60
Qy 104 TTTTAGCCAGTATTTCAAGTGTCTGTAGACAAGATGAATCATCTCCATTAATAGACGG 163
Db 61 TTTTAGCCAGTATTTCAAGTGTCTGTAGACAAGATGAATCATCTCCATTAATAGACGG 120
Qy 164 CAATGGACCTCACTATCATTTAGGGTTAACAGCCCAAGAACTTTCTTGTCAACAGAAC 223
Db 121 CAATGGACCTCACTATCATTTAGGGTTAACAGCCCAAGAACTTTCTTGTCAACAGAAC 180
Qy 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 283
Db 181 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 240
Qy 284 ATGGAGAAGAGAAAGTAACACCGAAATATCTTCCAGCACTTTAGAAAGGGGACCCCTG 343
Db 241 ATGGAGAAGAGAAAGTAACACCGAAATATCTTCCAGCACTTTAGAAAGGGGACCCCTG 300
Qy 344 ACTGTGTTAAGAGAGTGGGAGAACCCAGGCTGGGACAGAGTCTCACAGACTCT 403
Db 301 ACTGTGTTAAGAGAGTGGGAGAACCCAGGCTGGGACAGAGTCTCACAGACTCT 360
Qy 404 CTACGGAACAGCAGCACTGAGATTAGCAGACAGACACCATCTCTCTGCTGAAGTGACA 463
Db 361 CTACGGAACAGCAGCACTGAGATTAGCAGACAGACACCATCTCTCTGCTGAAGTGACA 420
Qy 464 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGACTCTAGA 523
Db 421 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGACTCTAGA 480
Qy 524 CTCAGTCACTCTGAGGCCCTGCTCAGGTGATATCCCAATCAAGAGCGGTGAG 583
Db 481 CTCAGTCACTCTGAGGCCCTGCTCAGGTGATATCCCAATCAAGAGCGGTGAG 540
Qy 584 GATCTTAAAGACCACTCAACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGG 643
Db 541 GATCTTAAAGACCACTCAACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGG 600
Qy 644 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAA 703
Db 601 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAA 660
Qy 704 TATTAATGTCGCTGAACAGCCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACT 763
Db 661 TATTAATGTCGCTGAACAGCCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACT 720
Qy 764 AAGATTCTCGGGCCCAAGCCGAGTGAAGTGAAGGAGAGATCTCTGAAGACAGCTAT 823
Db 721 AAGATTCTCGGGCCCAAGCCGAGTGAAGTGAAGGAGAGATCTCTGAAGACAGCTAT 780
Qy 824 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 883
Db 781 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 840
Qy 884 AAAAAATGAGAGTAGAGAAATCTGGAACTTCCAGCCTCTCAGAAACCTCTATAAAGGAT 943
Db 841 AAAAAATGAGAGTAGAGAAATCTGGAACTTCCAGCCTCTCAGAAACCTCTATAAAGGAT 900

Qy 944 CGAATGGCCAAAGTACCAAGCAGCTGTGTCCAAAACAAAGCAGCTCAACCAACTATACAAT 1003
Db 901 CGAATGGCCAAAGTACCAAGCAGCTGTGTCCAAAACAAAGCAGCTCAACCAACTATACAAT 960
Qy 1004 GAGCTGAAGAGCAGTGGTGGGAAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTG 1063
Db 961 GAGCTGAAGAGCAGTGGTGGGAAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTG 1020
Qy 1064 CCCCAGGTCCTGAGGTCTGCATCAACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1123
Db 1021 CCCCAGGTCCTGAGGTCTGCATCAACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1080
Qy 1124 AATAGCCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTACTCCAGGTTAAG 1183
Db 1081 AATAGCCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTACTCCAGGTTAAG 1140
Qy 1184 AGTAGGTTCAACAGCCTGTGCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCCAGT 1243
Db 1141 AGTAGGTTCAACAGCCTGTGCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCCAGT 1200
Qy 1244 CTTTCTGAAGTTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACC 1303
Db 1201 CTTTCTGAAGTTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACC 1260
Qy 1304 TCGTGGAAATGTCAAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG 1363
Db 1261 TCGTGGAAATGTCAAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG 1320
Qy 1364 TTTACATCAGTGTTCGTTGCTCTATTGCAACAAACACTCAGTCTAGGAACATAT 1423
Db 1321 TTTACATCAGTGTTCGTTGCTCTATTGCAACAAACACTCAGTCTAGGAACATAT 1380
Qy 1424 GCATCTTTACATGGAAGAAATCTATTGTAAGCTCACTTCAATCAACTCTTTAAATCTAAG 1483
Db 1381 GCATCTTTACATGGAAGAAATCTATTGTAAGCTCACTTCAATCAACTCTTTAAATCTAAG 1440
Qy 1484 GCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAT 1543
Db 1441 GCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAT 1500
Qy 1544 GAAAAAGAGAGATTTTGAGAGACAGCCAGCTTGCAAAATGCAAGGAGACCCCTCAC 1603
Db 1501 GAAAAAGAGAGATTTTGAGAGACAGCCAGCTTGCAAAATGCAAGGAGACCCCTCAC 1560
Qy 1604 AGCCAGGGTAGAAGATGCCCTTATTGCTAAGTGGGTGCTCTGGCTGCAAGTATGAA 1663
Db 1561 AGCCAGGGTAGAAGATGCCCTTATTGCTAAGTGGGTGCTCTGGCTGCAAGTATGAA 1620
Qy 1664 GCAAGGCTCTCTCTCAGCAGAGAGAGACAGCCAGCTGAACCAAGAGAGTGGG 1723
Db 1621 GCAAGGCTCTCTCTCAGCAGAGAGAGAGACAGCCAGCTGAACCAAGAGTGGG 1680
Qy 1724 ATCGCTGGCCACCCCTCACTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 1783
Db 1681 ATCGCTGGCCACCCCTCACTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 1740
Qy 1784 AAAATGTCAAAGCCCAATGGCTCTTGAAGCGAAATCAGCAAGCCCAAGTTTCTCTGAG 1843
Db 1741 AAAATGTCAAAGCCCAATGGCTCTTGAAGCGAAATCAGCAAGCCCAAGTTTCTCTGAG 1800
Qy 1844 GATGTCGATCTAGATCTGAAGAGTAAAGATCTTCTTCACTGAAGGAAAGAGAGCCGC 1903
Db 1801 GATGTCGATCTAGATCTGAAGAGTAAAGATCTTCTTCACTGAAGGAAAGAGAGCCGC 1860
Qy 1904 CCATTTCATGTAGAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAACTGTGTCC 1963
Db 1861 CCATTTCATGTAGAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAACTGTGTCC 1920
Qy 1964 CCACCTATCAGGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTCTGTGGTGGAAAGA 2023
Db 1921 CCACCTATCAGGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTCTGTGGTGGAAAGA 1980
Qy 2024 GTTGCAAGAGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTGGGA 2083

1981 GTTGCAGAAAGAAACAAAGTGGAAATGCGAAGCTTCTAAGAAAGATGGGAATGTGGGA 2040
2084 AAAACAACTCGGCAAAACAAAGATCTAAGAGAGAGACAGGAAAGAGAAAGGT 2143
2041 AAAACAACTCGGCAAAACAAAGATCTAAGAGAGAGACAGGAAAGAGAAAGGT 2100
2144 CATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAATGGTGAGACTCCGATGAAGAT 2203
2101 CATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAATGGTGAGACTCCGATGAAGAT 2160
2204 GATACAGCTTCTCAACACAACTCCCAAGAACCCAGTCTCTGAATTTGGTGGAGT 2263
2161 GATACAGCTTCTCAACACAACTCCCAAGAACCCAGTCTCTGAATTTGGTGGAGT 2220
2264 TTGTAGACAAACACTTGTCTGGAAGAAATTTCACTACTCAGAAATCCAGGATGTG 2323
2221 TTGTAGACAAACACTTGTCTGGAAGAAATTTCACTACTCAGAAATCCAGGATGTG 2280
2324 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAAAGAACAGATAAGAGAAAT 2383
2281 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAAAGAACAGATAAGAGAAAT 2340
2384 CGGTATTATGATGAGATGAGATGAGAGTGAAGATGCAAAATTCGAATGCTGGGCTTAAA 2443
2341 CGGTATTATGATGAGATGAGATGAGAGTGAAGATGCAAAATTCGAATGCTGGGCTTAAA 2400
2444 TTCATGTTAGTGTAGGAGGACCTGCGCTTTGTCAAAATGTGATGACACATAAGCAGTA 2503
2401 TTCATGTTAGTGTAGGAGGACCTGCGCTTTGTCAAAATGTGATGACACATAAGCAGTA 2460
2504 TCCAGCATGAAATGTAAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTTTAAAT 2563
2461 TCCAGCATGAAATGTAAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTTTAAAT 2520
2564 CAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2623
2521 CAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2580
2624 AATCTTCATTTAGCAGTATGATATGATGATGATGATGATGATGATGATGATGATGATG 2683
2581 AATCTTCATTTAGCAGTATGATATGATGATGATGATGATGATGATGATGATGATGATG 2640
2684 TATCCACCTGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2743
2641 TATCCACCTGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
2744 TAGATGATTAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2803
2701 TAGATGATTAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2760
2804 TTTAGGGCTTAAACATTTACACTGAAATGCACTTTAGTATTAAGGGGACAGTTTGTATAT 2863
2761 TTTAGGGCTTAAACATTTACACTGAAATGCACTTTAGTATTAAGGGGACAGTTTGTATAT 2820
2864 TTTAAATGAATACCAATTTAAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 2923
2821 TTTAAATGAATACCAATTTAAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 2880
2924 TAAATTTTTAGTTAAATTTTCTGCTGTATATATATATATATATATATATATATATATATAT 2983
2881 TAAATTTTTAGTTAAATTTTCTGCTGTATATATATATATATATATATATATATATATATAT 2940
2984 CTGCTCTTAAACTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 3043
2941 CTGCTCTTAAACTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 3000
3044 TTGAGGCAATTTGAAACCAACCTACCTCTTCCGCTGCTTAGAGATGCTGCTCTCC 3103
3001 TTGAGGCAATTTGAAACCAACCTACCTCTTCCGCTGCTTAGAGATGCTGCTCTCC 3060
3104 AAATAAGCTTTTGTATCTGCGAGTGAATTTACTGCTCTCCAAATGATGCTTTCTTCT 3163

3061 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAATGATGCTTTCTTTCT 3120
3164 GGTGATATCTGTCTTCTCATTAATTAAGAGCTGCAATATTTTAGTAAATACCTTTGGG 3223
3121 GGTGATATCTGTCTTCTCATTAATTAAGAGCTGCAATATTTTAGTAAATACCTTTGGG 3180
3224 ATCACTGTCCCCCATCTTCCGTGTTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAA 3283
3181 ATCACTGTCCCCCATCTTCCGTGTTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAA 3240
3284 AGAAGTGTCTTACACCACTTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGATGTC 3343
3241 AGAAGTGTCTTACACCACTTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGATGTC 3300
3344 CCCTTTTGGAGACACTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3403
3301 CCCTTTTGGAGACACTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3360
3404 ACAGTATCTCAGAGTGAAATTAACCAACTATAGAGCTTTTCTTGGAGATGATTTCTTA 3463
3361 ACAGTATCTCAGAGTGAAATTAACCAACTATAGAGCTTTTCTTGGAGATGATTTCTTA 3420
3464 GTCTTAAGCTTTGGGACACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTC 3523
3421 GTCTTAAGCTTTGGGACACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTC 3480
3524 AAATTTATGATGAGTGGAGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTGTACTGCAIT 3583
3481 AAATTTATGATGAGTGGAGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTGTACTGCAIT 3540
3584 TATAGAGATTTAGCTTTAATTAATTTTATAGAGATGTAACACATTTCTGCTTTCTTAGTCTTA 3643
3541 TATAGAGATTTAGCTTTAATTAATTTTATAGAGATGTAACACATTTCTGCTTTCTTAGTCTTA 3600
3644 CCTAGTCTGAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3703
3601 CCTAGTCTGAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3660
3704 AA 3705
3661 AA 3662

RESULT 2
HSM800164
LOCUS
DEFINITION
Homo sapiens mRNA; cDNA DKF2p586i1918 (from clone DKF2p586i1918);
complete cds.
ACCESSION
AL136911
VERSION
AL136911.1 GI:12053316
KEYWORDS
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3664)
Wambutt R., Heubner D., Mewes H.W., Gassenhuber J. and Wiemann S.
Submitted (15-JUL-2001) MIPF, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKF2p586i1918) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/
Location/Qualifiers
1. 3664
/organism="Homo sapiens"
/mol_type="mRNA"

```
/db_xref="taxon:9606"
/map="12"
/clone="DKF2p586i1918"
/issue_type="uterus"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
Dh10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1..3664
/gene="DKF2p586i1918"
CDS
94..2373
/gene="DKF2p586i1918"
/note="similarity to transcription factor SF3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66845.1"
/db_xref="GI:12053117"
/db_xref="SPTREMBL:Q9UHB6"
/translation="MESSFNRKQTSLRVTAKELSLVNKNKSAIVEIFSKYQKA
AETNMKRSNTENLSQHFRTGLTVLKKWENPGLGAESHTDSLRNSSTEIRHAD
HPFAEVTSHAAGKADQEOHPHRSRLRSPPEALVQGRPHIKDGEDLKHSTESKK
MENCLEGSHEVEKSEISENTDASGKIEKYNVPLNRLNMPEKGEPTOTKILRAQSR
ASGRKISENSYSLDLEIGPGLSSSTDSSEKVESRNLLELPLRSETSIKDRMAKYOA
AVSKOSSSNYNYNELKASGGEIKHKMEQKENVPPPEVCITHORGEKISANESLAV
RSTPAEDSRDSQVSEVOQPHPLPSDRASLSSESPKAWKFOAPARETCVE
CQKTYPMERHLANDQVHFISCFRSCYCNKLSLGTYSALHGRYICKPHFNQLFKSG
NYDEFGHPRHLDLWASKNENEIILERPAQLANARETPHSPGVEDAPIAKTVGLKASM
EAKASQOEKDEKPAETKKLRTAMPPTTELGSGLALBEGIKMSKPKWPPDEISKPE
VPEDVDLKLKLRSSLSIKERSRPTVAASFOSVSKPKTVSPPIRKGWMSYQSSEE
SVGGRVAKRKQVENAKSKNGNVKKTWQNKESGKETGKRSKEGHSLENNENLVEN
GADSDDDNSFLKQSPQEPKSLNSSFVDNTFAEFTTQNKQSDVELMEGEVVKEL
SVEEQIKRNYVDEDEDEE"
3637..3642
/polyA_signal
/gene="DKF2p586i1918"
3647
/polyA_site
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Query Match 97.5%; Score 3611; DB 9; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 44 GCGTAGGTAGAGCGCGGACCTGTGACAGGCTGTGTAGCAGCGCAGAGGAAAGCGCGC 103
DB 1 GCGCTAGGTAGAGCGCGGACCTGTGACAGGCTGTGTAGCAGCGCAGAGGAAAGCGCGC 60
QY 104 TTTTACCGAGGTATTTCACTGTCTGTAGACAGATGGAATCATCTCCATTTAATAGACGG 163
DB 61 TTTTACCGAGGTATTTCACTGTCTGTAGACAGATGGAATCATCTCCATTTAATAGACGG 120
QY 164 CAATGACCTCACTATCATTGAGGTAAACAGCCAAAGACTTTCTCTTGTCAACAGAAC 223
DB 121 CAATGACCTCACTATCATTGAGGTAAACAGCCAAAGAACTTTCTCTTGTCAACAGAAC 180
QY 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAAGTACAGAAAGCAGCTGAAGAAACAAAC 283
DB 181 AAGTCATCGGCTATTGTGGAATATTCTCCAAAGTACAGAAAGCAGCTGAAGAAACAAAC 240
QY 284 ATGGAGAAGAGAGAGTAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTG 343
DB 241 ATGGAGAAGAGAGAGTAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTG 300
QY 344 ACTGTGTTAAAGNAGAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT 403
DB 301 ACTGTGTTAAAGNAGAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT 360
QY 404 CTACGGAACAGAGCACTGAGATTAGGCAAGAGACCATCTCTCTGCTGAAGTGACA 463
DB 361 CTACGGAACAGAGCACTGAGATTAGGCAAGAGCAGACCATCTCTCTGCTGAAGTGACA 420
QY 464 AGCCACGCTGCTTCTGGAGCCAAAGGTGACCCAAAGAGAACAAATCCACCCAGACTCTAGA 523
DB 421 AGCCACGCTGCTTCTGGAGCCAAAGGTGACCCAAAGAGAACAAATCCACCCAGACTCTAGA 480
```

```
QY 524 CTCAGGTCACTCTCTGAAGCCCTCGTTCAAGGTCGATATATCCACATCAAGGACGGTGAG 583
DB 481 CTCAGGTCACTCTCTGAAGCCCTCGTTCAAGGTCGATATATCCACATCAAGGACGGTGAG 540
QY 584 GATCTTTAAAGACCACTCAACAGAAAGTAAATAAATGGAATAATTGTCTAGGAGAAATCCAGG 643
DB 541 GATCTTTAAAGACCACTCAACAGAAAGTAAATAAATGGAATAATTGTCTAGGAGAAATCCAGG 600
QY 644 CATGAAGTAGAAAAATCAGAAAAATCAGTGAATAACACAGATGCTCTCGGCAAAAAATAGAGAAA 703
DB 601 CATGAAGTAGAAAAATCAGAAAAATCAGTGAATAACACAGATGCTCTCGGCAAAAAATAGAGAAA 660
QY 704 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 763
DB 661 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 720
QY 764 AAGATTCTCCGGGCCCCAAACCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTAT 823
DB 721 AAGATTCTCCGGGCCCCAAACCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTAT 780
QY 824 TCTCTAGATGACCTGGAAAAATAGGCCCAAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 883
DB 781 TCTCTAGATGACCTGGAAAAATAGGCCCAAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 840
QY 884 AAAAAATCAGAGTAGACGAAATCTGGAACTTCCACGCTCTCAGAAAACTCTATAAAGGAT 943
DB 841 AAAAAATCAGAGTAGACGAAATCTGGAACTTCCACGCTCTCAGAAAACTCTATAAAGGAT 900
QY 944 CGAATGCCCAAGTACCAAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT 1003
DB 901 CGAATGCCCAAGTACCAAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT 960
QY 1004 GAGCTGAAAGCCAGTGTGTGGGAAATCAAAAATTCATAAATGGAGCAAAAGGAGAAATGTG 1063
DB 961 GAGCTGAAAGCCAGTGTGTGGGAAATCAAAAATTCATAAATGGAGCAAAAGGAGAAATGTG 1020
QY 1064 CCCCCAGGTCTCTGAGGTCGCATCACCCCTGCCGAAAGATGACTCCCGTGACTCCAGGTTAAG 1123
DB 1021 CCCCCAGGTCTCTGAGGTCGCATCACCCCTGCCGAAAGATGACTCCCGTGACTCCAGGTTAAG 1080
QY 1124 AATAGCCTGCGAGTCCGTTCCACCCCTGCCGAAAGATGACTCCCGTGACTCCAGGTTAAG 1183
DB 1081 AATAGCCTGCGAGTCCGTTCCACCCCTGCCGAAAGATGACTCCCGTGACTCCAGGTTAAG 1140
QY 1184 AGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCCTCCAGT 1243
DB 1141 AGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCCTCCAGT 1200
QY 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1303
DB 1201 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1260
QY 1304 TGGCTGGAATGTCAGAAAGACAGTCTATCCCAATGGAGGCTCTTTGGCCCAACCCAGCAGGTG 1363
DB 1261 TGGCTGGAATGTCAGAAAGACAGTCTATCCCAATGGAGGCTCTTTGGCCCAACCCAGCAGGTG 1320
QY 1364 TTTTCACTACGCTGCTTCCGTTGCTCCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1423
DB 1321 TTTTCACTACGCTGCTTCCGTTGCTCCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1380
QY 1424 GCATCTTTTATGGAAGAATCTATTGTAAGCCTCACTTTCAATCAACTCTTTTAAATCTAAG 1483
DB 1381 GCATCTTTTATGGAAGAATCTATTGTAAGCCTCACTTTCAATCAACTCTTTTAAATCTAAG 1440
QY 1484 GGCACCTATGATGAAGCTTTGGGCAAGACCCACACAGGATCTATGGGCAAGCAAAAAT 1543
DB 1441 GGCACCTATGATGAAGCTTTGGGCAAGACCCACACAGGATCTATGGGCAAGCAAAAAT 1500
QY 1544 GAAAAACGAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGGAGAGCCCTCAC 1603
DB 1501 GAAAAACGAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGGAGAGCCCTCAC 1560
```

QY 1604 AGCCAGGGGTAGAGATGCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAGTATGGAA 1663
DB 1561 AGCCAGGGGTAGAGATGCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAGTATGGAA 1620
QY 1664 GCCAAGGCTCTCTCAGCAGGAGAGAGACAAAGCCAGCTGAAACCAAGAGCTGAGG 1723
DB 1621 GCCAAGGCTCTCTCAGCAGGAGAGAGACAAAGCCAGCTGAAACCAAGAGCTGAGG 1680
QY 1724 ATCGGCTGCCACCCCTCAGCTGAACTGGAAGTTCAGGAGTGCCTTGGAGAGGGATC 1783
DB 1681 ATCGGCTGCCACCCCTCAGCTGAACTGGAAGTTCAGGAGTGCCTTGGAGAGGGATC 1740
QY 1784 AAAATGTCAGAGCCCAATGGCTCCTCAGAGCAGAAATCAGAACCCCGAGATTCCTGAG 1843
DB 1741 AAAATGTCAGAGCCCAATGGCTCCTCAGAGCAGAAATCAGAACCCCGAGATTCCTGAG 1800
QY 1844 GATGTCGATCTAGATCTGAAGAAGCTTAAGCATCTTCTTCACTGAAGGAAAGAGCCGC 1903
DB 1801 GATGTCGATCTAGATCTGAAGAAGCTTAAGCATCTTCTTCACTGAAGGAAAGAGCCGC 1860
QY 1904 CATTCACGTGAGAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAACTGTGTCC 1963
DB 1861 CATTCACGTGAGAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAACTGTGTCC 1920
QY 1964 CCACCTATCAGGAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGTGGAGA 2023
DB 1921 CCACCTATCAGGAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGTGGAGA 1980
QY 2024 GTTGAG 2083
DB 1981 GTTGAG 2040
QY 2084 AAAACAACTGCGCAAAAACAAAGAACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2143
DB 2041 AAAACAACTGCGCAAAAACAAAGAACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2144 CATAGTTTGGAGATGGAGATGAGATCTTGTAGAAATGTTGAGACTCCGATGAAGAT 2203
DB 2101 CATAGTTTGGAGATGGAGATGAGATCTTGTAGAAATGTTGAGACTCCGATGAAGAT 2160
QY 2204 GATAACAGCTTCTCAACAAACAACTCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAG 2263
DB 2161 GATAACAGCTTCTCAACAAACAACTCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2264 TTTGTAGACAAACCTTTGCTGAAGAACTTCACTACTCAGAAATCCAGAGTGTG 2323
DB 2221 TTTGTAGACAAACCTTTGCTGAAGAACTTCACTACTCAGAAATCCAGAGTGTG 2280
QY 2324 GAACTCTGGAGGAGAGAGTGTGCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2383
DB 2281 GAACTCTGGAGGAGAGAGTGTGCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2340
QY 2384 CGGTATTATGATGAGATGAGATGAGAGTGAACAAATGCAATGATGCTGGGCTTAAA 2443
DB 2341 CGGTATTATGATGAGATGAGATGAGAGTGAACAAATGCAATGATGCTGGGCTTAAA 2400
QY 2444 TTCATGTTAGTGTAGGAGGACCTGCCCTTTGTCAAAATGTCATGACATAAGCAGGTA 2503
DB 2401 TTCATGTTAGTGTAGGAGGACCTGCCCTTTGTCAAAATGTCATGACATAAGCAGGTA 2460
QY 2504 TCCCAGCATGAAATGTAAATTTACTTGAAGTAACCTTTGGAAAAGAAATTCCTTTTAAAT 2563
DB 2461 TCCCAGCATGAAATGTAAATTTACTTGAAGTAACCTTTGGAAAAGAAATTCCTTTTAAAT 2520
QY 2564 CAAAAACAAAAACAAAAACAAAAACAACTTCTAAATCTAGAGATACTTACTTA 2623
DB 2521 CAAAAACAAAAACAAAAACAAAAACAACTTCTAAATCTAGAGATACTTACTTA 2580
QY 2624 AATCTTCAATTTAGCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2683
DB 2581 AATCTTCAATTTAGCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2640
QY 2684 TATTCCACCTGATAATGCCAGATTTCTACTGTATTCCTCAAAAGGCAATATTAGGTAGA 2743

RESULT 3

DB 2641 TATTCCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAGGTAGA 2700
QY 2744 TAGATGATTAGTAGTATATTGTTTACACACTATTGGAATTTAGAGAACATACAGAGAGAA 2803
DB 2701 TAGATGATTAGTAGTATATTGTTTACACACTATTGGAATTTAGAGAACATACAGAGAGAA 2760
QY 2804 TTTAGGGCTTTAAACATTACGACTGAATGCACTTTAGTATATAAGGGGACAGTTTGTATAT 2863
DB 2761 TTTAGGGCTTTAAACATTACGACTGAATGCACTTTAGTATATAAGGGGACAGTTTGTATAT 2820
QY 2864 TTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTTAAGAGATATTATTAGTCTT 2923
DB 2821 TTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTTAAGAGATATTATTAGTCTT 2880
QY 2924 TAAATTTTGTAGTAAATTTCTGCTGATATATATAGGAAATTTACTACTTTTATGTC 2983
DB 2881 TAAATTTTGTAGTAAATTTCTGCTGATATATATAGGAAATTTACTACTTTTATGTC 2940
QY 2984 CTGCTCTCTAAACTACATCTGAACTCGAGTCTCTGAGGTATATAACACAGAGACACTTT 3043
DB 2941 CTGCTCTCTAAACTACATCTGAACTCGAGTCTCTGAGGTATATAACACAGAGACACTTT 3000
QY 3044 TTGAGGCAATTTGAAAACCAACCTACACTCTTCGGTCTTTAGAGAGATCTGCTGTCTCCC 3103
DB 3001 TTGAGGCAATTTGAAAACCAACCTACACTCTTCGGTCTTTAGAGAGATCTGCTGTCTCCC 3060
QY 3104 AATATAGCTTTTCTATCTGACAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCT 3163
DB 3061 AATATAGCTTTTCTATCTGACAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCT 3120
QY 3164 GGTGATATCTGCTCTCTAATAATTTACTGAAAGCTGCAATATTTTTAGTAATACCTTCGGG 3223
DB 3121 GGTGATATCTGCTCTCTAATAATTTACTGAAAGCTGCAATATTTTTAGTAATACCTTCGGG 3180
QY 3224 ATCACTGTCCCCATCTTCGGTCTTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAGAA 3283
DB 3181 ATCACTGTCCCCATCTTCGGTCTTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAGAA 3240
QY 3284 AGAAGTGTCTTACACACTGAGCTCAGACCTCTAAACCTGTATTTCCTTATGATGTC 3343
DB 3241 AGAAGTGTCTTACACACTGAGCTCAGACCTCTAAACCTGTATTTCCTTATGATGTC 3300
QY 3344 CCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTGATTTTATC 3403
DB 3301 CCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTGATTTTATC 3360
QY 3404 ACAGTATTTCTAGGGTGAATTTAAACCACTATAGGCTTTTCTTTGGGATGATTTCTTA 3463
DB 3361 ACAGTATTTCTAGGGTGAATTTAAACCACTATAGGCTTTTCTTTGGGATGATTTCTTA 3420
QY 3464 GTCTTAAGTGTGGGACATTTAAACCTTGAATACATTTTGTGTACACAGTGTGATTTCC 3523
DB 3421 GTCTTAAGTGTGGGACATTTAAACCTTGAATACATTTTGTGTACACAGTGTGATTTCC 3480
QY 3524 AATTTGATGGATGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGTACTGCAAT 3583
DB 3481 AATTTGATGGATGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGTACTGCAAT 3540
QY 3584 TATAGAGATTTAGCTTTAATATTTTGTAGAGTGTAAACATTTCTCTTTCTTAGTCTTA 3643
DB 3541 TATAGAGATTTAGCTTTAATATTTTGTAGAGTGTAAACATTTCTCTTTCTTAGTCTTA 3600
QY 3644 CCTAGTCTGAACATTTTATTCATATAAGATTTTAAATTTGAAAAAATAAAAAA 3703
DB 3601 CCTAGTCTGAACATTTTATTCATATAAGATTTTAAATTTGAAAAAATAAAAAA 3660
QY 3704 AA 3705
DB 3661 AA 3662

AF198454
LOCUS AF198454 3655 bp mRNA linear PRI 10-JAN-2000
DEFINITION Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA, complete cds.
ACCESSION AF198454
VERSION AF198454.1 GI:6685006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3655)
Maul, R.S. and Chang, D.D.
EPLIN, epithelial protein lost in neoplasm
ONCOGENE 18 (54), 7838-7841 (1999)
MEDLINE 20087188
PUBMED 10618726
2 (bases 1 to 3655)
Maul, R.S. and Chang, D.D.
Direct Submission
TITLE Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
JOURNAL
FEATURES
source
1..3655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
1..3655
/gene="EPLIN"
102..2381
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAP23755.1"
/db_xref="GI:6685007"
translation="MESSPFNRQWTSLSLRVTAKELSLVKNKSSAIVEIFSKYKA
AEETNMKKSNTENSLQHFQKGLTVLKKKNWENPGLGAEHSHTDSLRNSSTIEH
HPEAVTSHAASGAKADQEEQIHPRSLRPPALVQGRYPHFKDGEDLKHSTESKK
MENCLGSRHEVEKSEISENDASGKIERKINFLNRLKMFKEGPTQTKLRAQSR
ASRKKSENYSYLDLLEIGQLSSSTFSEKVESRNLELPLRSLSIKDRMAYQA
AVQSSTSYNTHLKASGGEIKIHQMEQENVPPEVCIHQEGEKISANENSLAV
RSTPAEDSDRSQVSEVQVPHKPLSPDRASSLSSESPKMKKFOAPARETCVE
COKTVYPMERLLANQOVFIHSCPRCSYCNKLISLGTVASLHGRIYCKPHNQLFKSG
NYDEGFHRPHKOLWASKNENELBERPAQLANARETPSPGVEDAPIKQVLAASM
EKASSQEDKEDPAETKURIAWPPPTLGGSSALEEGIKMSKPKWPPDEISEKPE
VPDVLDDLKLRSSSLKERSRPFVVAASFQTSVKSPKTVSPPIKGMSESEEE
SVGRVAERQVENAKSKKNVGNKTVNQNKESKGETGRSKEGHSLEWENENLVEN
GADSDDDNSFLKQSPQEPKSLNWSFVNDTFAEEFTTQNKQSDVELWEGEVKEL
SVSEQIKRNYDEDEDEE"
polyA_signal 3632..3637
/gene="EPLIN"
BASE COUNT 1197 a 737 c 810 g 911 t
ORIGIN
Query Match 95.8%; Score 3548; DB 9; Length 3655;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 92 AGGAAGCGCGCTTTTAGCCAGGTATTTACAGTGTCTGTAGACAAGATGGATCATCTCCA 151
DB 57 AGGAAGCGCGCTTTTAGCCAGGTATTTACAGTGTCTGTAGACAAGATGGATCATCTCCA 116
QY 152 TTAAATAGCGGCAATGGAGCTCACTATCATTTAGGGGTAAACGCCAAAGAACTTTCTCTT 211
DB 117 TTAAATAGCGGCAATGGAGCTCACTATCATTTAGGGGTAAACGCCAAAGAACTTTCTCTT 176
QY 212 GTCAACAGAACAGTATCGGCTATTGTGGAATATTTCCAGTACCGAAGAGAGCT 271
DB 177 GTCAACAGAACAGTATCGGCTATTGTGGAATATTTCCAGTACCGAAGAGAGCT 236
QY 272 GAAGAAACAAACATGGAGAGAGAGAGAGTAACACGCAAAATCTCTCCGAGCACTTTAGA 331

237 GAAGAAACAAACATGGAGAGAGAGAGTAACACGCAAAATCTCTCCGAGCACTTTAGA 296
QY 332 AAGGGGACCCCTGACTGTGTAAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGTCT 391
DB 297 AAGGGGACCCCTGACTGTGTAAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGTCT 356
QY 392 CACACAGACTCTCTACGGAAACAGCAGCACTGAGATTAGGCACAGAGCAGACCATCTCTCT 451
DB 357 CACACAGACTCTCTACGGAAACAGCAGCACTGAGATTAGGCACAGAGCAGACCATCTCTCT 416
QY 452 GCTGAAGTGAACAGCAGCAGCTGCTTTCTGGAGCAAAAGCTGACCAAGAGAAACAAATCCAC 511
DB 417 GCTGAAGTGAACAGCAGCAGCTGCTTTCTGGAGCAAAAGCTGACCAAGAGAAACAAATCCAC 476
QY 512 CCAGATCTAGACTCAGCTCAGCTCCTCTGAGCCCTCGTTGAGGGTCTGATATCCCCACATC 571
DB 477 CCAGATCTAGACTCAGCTCAGCTCCTCTGAGCCCTCGTTGAGGGTCTGATATCCCCACATC 536
QY 572 AAGGACGGTGAAGATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGAAAAATGTGCTA 631
DB 537 AAGGACGGTGAAGATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGAAAAATGTGCTA 596
QY 632 GGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGC 691
DB 597 GGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGC 656
QY 692 AAAATAGAAATATAATGTTCCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGA 751
DB 657 AAAATAGAAATATAATGTTCCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGA 716
QY 752 CCAACTCAAACTAAGATTTCTCGGGCCCAAGCGCAAGTGAAGTGAAGGAGATCTCT 811
DB 717 CCAACTCAAACTAAGATTTCTCGGGCCCAAGCGCAAGTGAAGTGAAGGAGATCTCT 776
QY 812 GAAAAACAGCTATTTCTTAGATGACCTTGAATATAGCCAGGTCAGTTGTCATCTTTCTACA 871
DB 777 GAAAAACAGCTATTTCTTAGATGACCTTGAATATAGCCAGGTCAGTTGTCATCTTTCTACA 836
QY 872 TTGACTCGGAAAAATAGAGATGAGCAAAATCTGGAATCTTCAACGCTCTTCAAGAAC 931
DB 837 TTGACTCGGAAAAATAGAGATGAGCAAAATCTGGAATCTTCAACGCTCTTCAAGAAC 896
QY 932 TCTATAAGAGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACAAAGCAGCTCAAC 991
DB 897 TCTATAAGAGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACAAAGCAGCTCAAC 956
QY 992 AACTATACAAATGAGCTGAAAGCCAGTGTGCGGAAATCNAATTCATAAATCGAGCAA 1051
DB 957 AACTATACAAATGAGCTGAAAGCCAGTGTGCGGAAATCNAATTCATAAATCGAGCAA 1016
QY 1052 AAGGAAATGTGCCCCAGGCTCTGAGGTCTGCAATCAACCATCAGGAAGGGGAAAAAGATT 1111
DB 1017 AAGGAAATGTGCCCCAGGCTCTGAGGTCTGCAATCAACCATCAGGAAGGGGAAAAAGATT 1076
QY 1112 TCTGCAATAGAAATAGCTCGGAGTCTGTTCCACCCCTGCGGAGATGATCTCCCGTAC 1171
DB 1077 TCTGCAATAGAAATAGCTCGGAGTCTGTTCCACCCCTGCGGAGATGATCTCCCGTAC 1136
QY 1172 TCCAGGTTAAGAGTGAAGTTCAACAGCTGTCCATCCAGGCCACTAAGTCCAGATTC 1231
DB 1137 TCCAGGTTAAGAGTGAAGTTCAACAGCTGTCCATCCAGGCCACTAAGTCCAGATTC 1196
QY 1232 AGAGCTCCAGTCTTTCTGAAAAGTTCTCTCCCAAGCAATGAAAGAGTTTTCAGSCACT 1291
DB 1197 AGAGCTCCAGTCTTTCTGAAAAGTTCTCTCCCAAGCAATGAAAGAGTTTTCAGSCACT 1256
QY 1292 GCAAGAGAGACTCGTGGAAATGTGAGAGAGTCTATCCAAATGGAGCGTCTCTTGGCC 1351
DB 1257 GCAAGAGAGACTCGTGGAAATGTGAGAGAGTCTATCCAAATGGAGCGTCTCTTGGCC 1316
QY 1352 AACAGAGAGTGTTCACATCAGCTGCTTCCGTTGCTCTTATTCGAAACAAACACTCAGT 1411

Db 1317 AACGACGAGGTGTTTCCATCATCAGCTGCTTCCGTTGCTCCTATTGCAACAAACAACTCAGT 1376
 Qy 1412 CTAGGAACATATGCATCTTATCATGCAAGAACTCTATTGTAAAGCCTCACCTTCAATCACTC 1471
 Db 1377 CTAGGAACATATGCATCTTATCATGCAAGAACTCTATTGTAAAGCCTCACCTTCAATCACTC 1436
 Qy 1472 TTTAAATCTAAAGGCAACTATGATGAAGGCTTTGGGCACAGACCACACAAAGGATCTATGG 1531
 Db 1437 TTTAAATCTAAAGGCAACTATGATGAAGGCTTTGGGCACAGACCACACAAAGGATCTATGG 1496
 Qy 1532 GCAAGCAAAATGAAACGAAAGATTTTGGAGAGACCAAGCCAGCTTGCAAAATGCAAGG 1591
 Db 1497 GCAAGCAAAATGAAACGAAAGATTTTGGAGAGACCAAGCCAGCTTGCAAAATGCAAGG 1556
 Qy 1592 GAGACCCCTCAGACCCAGGGGTAGAGATGCCCTTATTGCTTAAGTGGGTGCTCGCT 1651
 Db 1557 GAGACCCCTCAGACCCAGGGGTAGAGATGCCCTTATTGCTTAAGTGGGTGCTCGCT 1616
 Qy 1652 GCAAGTATGGAAGCCAAAGCCCTCTCAGCAGGAGGAAGAGCAAGCCAGCTGAAACC 1711
 Db 1617 GCAAGTATGGAAGCCAAAGCCCTCTCAGCAGGAGGAAGAGCAAGCCAGCTGAAACC 1676
 Qy 1712 AAGAAGCTGAGGATCGCCTGGCCACCCCTCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTG 1771
 Db 1677 AAGAAGCTGAGGATCGCCTGGCCACCCCTCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTG 1736
 Qy 1772 GAGGAAGGATCAAAATGTCAAAGCCAAATGGCCTCTGGAAGACGAAATCAGCAAGGCC 1831
 Db 1737 GAGGAAGGATCAAAATGTCAAAGCCAAATGGCCTCTGGAAGACGAAATCAGCAAGGCC 1796
 Qy 1832 GAAGTTCCTGAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTCACTGAAG 1891
 Db 1797 GAAGTTCCTGAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTCACTGAAG 1856
 Qy 1892 GAAAGAGCGGCCCACTTCACTGAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCA 1951
 Db 1857 GAAAGAGCGGCCCACTTCACTGAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCA 1916
 Qy 1952 AAAATGTTCTCCCACTATCAGGAAGAGCTGGAGCATGTTCAGACAGAGTGAAGAGTCT 2011
 Db 1917 AAAATGTTCTCCCACTATCAGGAAGAGCTGGAGCATGTTCAGACAGAGTGAAGAGTCT 1976
 Qy 2012 GTGGGTGGAAGAGTTCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAAAGAT 2071
 Db 1977 GTGGGTGGAAGAGTTCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAAAGAT 2036
 Qy 2072 GGGAAATGTGGGAAAACAACTTGGCAAAACAAAGAACTCTTAAGAGAGACAGGGAGAGA 2131
 Db 2037 GGGAAATGTGGGAAAACAACTTGGCAAAACAAAGAACTCTTAAGAGAGACAGGGAGAGA 2096
 Qy 2132 AGTAAGGAAGGTCTATGTTTGGAGATGGAGAAATGAGAATCTTGTAGAAAATGGTGCAGAC 2191
 Db 2097 AGTAAGGAAGGTCTATGTTTGGAGATGGAGAAATGAGAATCTTGTAGAAAATGGTGCAGAC 2156
 Qy 2192 TCCGATGAAGATGATAACAGCTTCTTCAAAACAAATCTCACAAGAACCCCAAGTCTCTG 2251
 Db 2157 TCCGATGAAGATGATAACAGCTTCTTCAAAACAAATCTCACAAGAACCCCAAGTCTCTG 2216
 Qy 2252 AATTGGTGGAGTTTGTAGACAAACACCTTGTGCTGAAGAAATTCACCTCAGAAATCAGAAA 2311
 Db 2217 AATTGGTGGAGTTTGTAGACAAACACCTTGTGCTGAAGAAATTCACCTCAGAAATCAGAAA 2276
 Qy 2312 TCCAGGATGTGGAATCTTGGAGGGAGGAAGTGGTCAAAAGAGCTCTCTGTGGAAGACAG 2371
 Db 2277 TCCAGGATGTGGAATCTTGGAGGGAGGAAGTGGTCAAAAGAGCTCTCTGTGGAAGACAG 2336
 Qy 2372 ATAAAGAGAAATCGGTATTTATGATGAGGATGAGGATGAAGAGTGCAAAATGCAATGATG 2431
 Db 2337 ATAAAGAGAAATCGGTATTTATGATGAGGATGAGGATGAAGAGTGCAAAATGCAATGATG 2396
 Qy 2432 CTGGCCCTTAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGTGATGCA 2491
 Db 2397 CTGGCCCTTAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGTGATGCA 2456

Qy 2492 CATAAAGCAGGTATCCAGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATT 2551
 Db 2457 CATAAAGCAGGTATCCAGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATT 2516
 Qy 2552 CCTTCTTAAATCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2611
 Db 2517 CCTTCTTAAATCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2576
 Qy 2612 TAACCTTTACTTTAAATTTCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATG 2671
 Db 2577 TAACCTTTACTTTAAATTTCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATG 2636
 Qy 2672 TAACCTGGGAAATTTCCACTGATTAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAA 2731
 Db 2637 TAACCTGGGAAATTTCCACTGATTAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAA 2696
 Qy 2732 TATTAAAGGTAGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2791
 Db 2697 TATTAAAGGTAGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2756
 Qy 2792 ATACAGAAGGAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTTAGTATTAAGGCA 2851
 Db 2757 ATACAGAAGGAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTTAGTATTAAGGCA 2816
 Qy 2852 CAGTTTGTATTTTAAATGAATACCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2911
 Db 2817 CAGTTTGTATTTTAAATGAATACCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2876
 Qy 2912 TTATTTAGTCTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2971
 Db 2877 TTATTTAGTCTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2936
 Qy 2972 CTACTTTATGTCGCTCTCTTAACCTACCTGAACTGCACTGCACTGCACTGCACTGCACTGCA 3031
 Db 2937 CTACTTTATGTCGCTCTCTTAACCTACCTGAACTGCACTGCACTGCACTGCACTGCACTGCA 2996
 Qy 3032 ACAGAGCACTTTTGGGCAATTTGAAACCAACCTACACTCTTTCGCTGCTTACGAGAGAT 3091
 Db 2997 ACAGAGCACTTTTGGGCAATTTGAAACCAACCTACACTCTTTCGCTGCTTACGAGAGAT 3056
 Qy 3092 CTGCTGTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAAATTTACTGTACTTCCAAATGATT 3151
 Db 3057 CTGCTGTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAAATTTACTGTACTTCCAAATGATT 3116
 Qy 3152 GCCTTCTTTCTGGTGATCTGCTCTTCTCAATTTACTGAAAGCTGCAATTTTACT 3211
 Db 3117 GCCTTCTTTCTGGTGATCTGCTCTTCTCAATTTACTGAAAGCTGCAATTTTACT 3176
 Qy 3212 AATACCTTCGGGATCACTGTCCTCCCTCTTCCGCTGTTAGAGCAAGTGAAGAGTTTAAAG 3271
 Db 3177 AATACCTTCGGGATCACTGTCCTCCCTCTTCCGCTGTTAGAGCAAGTGAAGAGTTTAAAG 3236
 Qy 3272 GAGGAAGAAAGAACTGCTTTACACCTTTCAGCTCAGACCTCTTAAACCCCTGTATTTTC 3331
 Db 3237 GAGGAAGAAAGAACTGCTTTACACCTTTCAGCTCAGACCTCTTAAACCCCTGTATTTTC 3296
 Qy 3332 CCTTATGATCTCCCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATA 3391
 Db 3297 CCTTATGATCTCCCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATA 3356
 Qy 3392 TTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAGGCTTTTCTTCTGG 3451
 Db 3357 TTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAGGCTTTTCTTCTGG 3416
 Qy 3452 GATGATTTTCTAGCTTTAGGTTTGGGCACTTATAAATTTAGTACATTTGTTGTACAC 3511
 Db 3417 GATGATTTTCTAGCTTTAGGTTTGGGCACTTATAAATTTAGTACATTTGTTGTACAC 3476
 Qy 3512 AGTTGATTTTCCAAATTTGTATGATGGAGGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCT 3571
 Db 3477 AGTTGATTTTCCAAATTTGTATGATGGAGGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCT 3536

QY 3572 TTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAAACATTTCTGCT 3631
Db 3537 TTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAAACATTTCTGCT 3596
QY 3632 TTCTTAGCTTACTAGCTGGAACATTTTATTCATTAAGAGATTTTAAATTAATTTG 3590
Db 3597 TTCTTAGCTTACTAGCTGGAACATTTTATTCATTAAGAGATTTTAAATTAATTTG 3655

RESULT 4
AK000335 3277 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20328 fis, clone HEP10039.
DEFINITION AK000335
VERSION AK000335.1 GI:7020350
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3277)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission

TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
JOURNAL University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,
AUTHORS Minato-Ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source Location/Qualifiers
1. .3277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP10039"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/notes="cloning vector pME18SFL3"
179..1978
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91092.1"
/db_xref="GI:7020351"
/translations="MNCIGESRHEVEKSEISENTDASQIKYVPLNRLKMFEGK
EPQTILRAQSASGRKISENSYSLLDLGPGGLSSSTFDSKESNRRLLEPRL
SETIKDRWAKYQAAVKSSSTNYNELKASGEIKIHMEQKNVPPGVCITHQ
EGKISANENSLAVRTPAEDSDRSQVKSQVPPVHPKPLSPDRASSLSSESPKA
MKFQAPARETCVCKQTVPMERLLANQVPHISCFRCYCNKLISLGTASLHGR
YCKPHNOLFKSGNYDEGHRPHKDLWASKNEEILERPQALANARETPHSQVE
DAPKAVGLAASMEAKSQQEKDPAETKRLIARWPPTTELSSGSALBEGIKMS
KPKWPDEISKEPVEDVLDLKKURSSLSUKERSRPFTVAASQTSVSKPXTVSP
PIKQMSMEQSESVGGVAVKQVENAKSKNGNVGKTTWQNKSEKGETGRSKE
GHSLENNENLNGADSDSDNSFLKQSQPOEPKPSLWSSSFVDNTFABEFTTQOKS
QDVELWEGEVVKLSVEEQIKNRYDEDEDESE"
BASE COUNT 1089 a 651 c 703 g 834 t
ORIGIN

Query Match 84.0%; Score 3114; DB 9; Length 3277;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 3264; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 439 AGACCATCTCTGCTGAAGTGAACAGCAGCTGCTTCTTGAGCCAAAGCTGACCAAGA 498
Db 1 AGACCATCTCTGCTGAAGTGAACAGCAGCTGCTTCTTGAGCCAAAGCTGACCAAGA 60
QY 439 AGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCTGTTCAAGGTCG 558
Db 61 AGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCTGTTCAAGGTCG 120
QY 559 ATATCCCCACATCAAGGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGCTAAATAAT 618
Db 121 ATATCCCCACATCAAGGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGCTAAATAAT 180
QY 619 GGAAATTTGCTAGAGAAATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAAACAC 678
Db 181 GGAAATTTGCTAGAGAAATCCAGGCATGAAGTAGAATAATCAGTAATCAGTGAAACAC 240
QY 679 AGATGCTTCGGGCAAAATAGAGAAATATAATTTCCGCTGAACAGGCTTAAGATGATGTT 738
Db 241 AGATGCTTCGGGCAAAATAGAGAAATATAATTTCCGCTGAACAGGCTTAAGATGATGTT 300
QY 739 TGAGAAAGGTGAACCAACTCAACTTAAGATTTCTCGGGCCCAAGCCGAAAGTGAAGTGG 798
Db 301 TGAGAAAGGTGAACCAACTCAACTTAAGATTTCTCGGGCCCAAGCCGAAAGTGAAGTGG 360
QY 799 AAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAATAGCCAGGTGACGTT 858
Db 361 AAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAATAGCCAGGTGACGTT 420
QY 859 GTCATCTTTTACATTTGATCTCGAGAGAAATAGAGTAGAGAGAAATCTGGAACCTTCCACG 918
Db 421 GTCATCTTTTACATTTGATCTCGAGAGAAATAGAGTAGAGAGAAATCTGGAACCTTCCACG 480
QY 919 CCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACA 978
Db 481 CCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACA 540
QY 979 AAGCAGCTCAACCACTATACAAATAGCTCAAGAGCCAGTGGTGGGAAATCAAAATTC 1038
Db 541 AAGCAGCTCAACCACTATACAAATAGCTCAAGAGCCAGTGGTGGGAAATCAAAATTC 600
QY 1039 TAAATGGAGCAAAAGGAGAAATGTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTCAGGA 1098
Db 601 TAAATGGAGCAAAAGGAGAAATGTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTCAGGA 660
QY 1099 AGGGGAAAAGATTTTCTGCAAAATGAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGA 1158
Db 661 AGGGGAAAAGATTTTCTGCAAAATGAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGA 720
QY 1159 TGACTCCCGTGACTCCCGAGGTTAAGAGTAGGTTCAACAGGCTGTCCATCCCAAGCCACT 1218
Db 721 TGACTCCCGTGACTCCCGAGGTTAAGAGTAGGTTCAACAGGCTGTCCATCCCAAGCCACT 780
QY 1219 AAGTCCAGATTTCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCCTCCCAAGCAATGAAGA 1278
Db 781 AAGTCCAGATTTCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCCTCCCAAGCAATGAAGA 840
QY 1279 GTTTCAGCACCTCTCAAGAGAGACCTCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGA 1338
Db 841 GTTTCAGCACCTCTCAAGAGAGACCTCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGA 900
QY 1339 CGCTCTCTTGGCCAAACAGCAGGTTTTCACATCAGTGTCTCCGTTGCTCTCTATTGCAA 1398
Db 901 CGCTCTCTTGGCCAAACAGCAGGTTTTCACATCAGTGTCTCCGTTGCTCTCTATTGCAA 960
QY 1399 CAAACAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTTATGTAAGCTCA 1458
Db 961 CAAACAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTTATGTAAGCTCA 1020
QY 1459 CTTCAATCAACTCTTTTAAATCTAAGGCAACTATGATGAAGGCTTTTGGGCACAGCCACA 1518
Db 1021 CTTCAATCAACTCTTTTAAATCTAAGGCAACTATGATGAAGGCTTTTGGGCACAGCCACA 1080

QY 1519 CAAGGATCTATGGGCAAGCAGAAATGAAACGAGAGATTTTGGAGAGACCAAGCCAGCT 1578
Db 1081 CAAGGATCTATGGGCAAGCAGAAATGAAACGAGAGATTTTGGAGAGACCAAGCCAGCT 1140
QY 1579 TGCAAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAAGGT 1638
Db 1141 TGCAAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAAGGT 1200
QY 1639 GGGTGTCTGGTGTGCAAGTATGGAAGCCAGGCCCTCTCTCAGCAGGAGGAAGAAAGACAA 1698
Db 1201 GGGTGTCTGGTGTGCAAGTATGGAAGCCAGGCCCTCTCTCAGCAGGAGGAAGAAAGACAA 1260
QY 1699 GCCAGCTGAAACCAAGAGCTGAGGATGCCCTGGCCACCCCTCAGTAACTTGGAGATTC 1758
Db 1261 GCCAGCTGAAACCAAGAGCTGAGGATGCCCTGGCCACCCCTCAGTAACTTGGAGATTC 1320
QY 1759 AGGAAGTGGCTGGAGGAGAGGATCAAAATGTCAAAGCCCAAAATGSCCTCTCGAAGACGA 1818
Db 1321 AGGAAGTGGCTGGAGGAGAGGATCAAAATGTCAAAGCCCAAAATGSCCTCTCGAAGACGA 1380
QY 1819 AATCAGCAAGCCGGAAGTCTCTGAGGATGTGATCTAGATCTGAAGAACTTAAGACGATC 1878
Db 1381 AATCAGCAAGCCGGAAGTCTCTGAGGATGTGATCTAGATCTGAAGAACTTAAGACGATC 1440
QY 1879 TTCTTCACTGAAGGAAGAGCCGCCATTTCACTGTAGCAGGTTCAATTTCAAAGACCTTC 1938
Db 1441 TTCTTCACTGAAGGAAGAGCCGCCATTTCACTGTAGCAGGTTCAATTTCAAAGACCTTC 1500
QY 1939 TGTCAAGACCCCAAAACTGTGTCCCACTATCAGGAAGGCTGGAGCATGTCAAGCA 1998
Db 1501 TGTCAAGACCCCAAAACTGTGTCCCACTATCAGGAAGGCTGGAGCATGTCAAGCA 1560
QY 1999 GAGTGAAGAGTCTGTGGTGGAGAGTGTGCAAGAGGAAACAAAGTGGAAATGCCCAAGGC 2058
Db 1561 GAGTGAAGAGTCTGTGGTGGAGAGTGTGCAAGAGGAAACAAAGTGGAAATGCCCAAGGC 1620
QY 2059 TTCTAAGAGAGATGGGAATGTGGGAAAAACACCTGGCAAAAAACAAAGATCTAAAGGAGA 2118
Db 1621 TTCTAAGAGAGATGGGAATGTGGGAAAAACACCTGGCAAAAAACAAAGATCTAAAGGAGA 1680
QY 2119 GACAGGAGAGAGATGAAGAGGTCATAGTTTGGAGATGGAGATGAAGATCTTTGTAGA 2178
Db 1681 GACAGGAGAGAGATGAAGAGGTCATAGTTTGGAGATGGAGATGAAGATCTTTGTAGA 1740
QY 2179 AAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCAACACACATCTCCACAGA 2238
Db 1741 AAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCAACACACATCTCCACAGA 1800
QY 2239 ACCCAAGTCTCTGAATTTGTCGAGTTTCTAGACAAACACCTTTGCTGAAGAATTCACCTAC 2298
Db 1801 ACCCAAGTCTCTGAATTTGTCGAGTTTCTAGACAAACACCTTTGCTGAAGAATTCACCTAC 1860
QY 2299 TCAGAAATCAGAAATCCAGGATGTGAATCTGGGAGGAGAGTGGTCAAAGAGCTCTC 2358
Db 1861 TCAGAAATCAGAAATCCAGGATGTGAATCTGGGAGGAGAGTGGTCAAAGAGCTCTC 1920
QY 2359 TGTGGAAGACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 2418
Db 1921 TGTGGAAGACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 1980
QY 2419 AAATGCAATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCAGCTGCCCTTTGTC 2478
Db 1981 AAATGCAATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCAGCTGCCCTTTGTC 2040
QY 2479 AAAATGTGATGCACATPAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACT 2538
Db 2041 AAAATGTGATGCACATPAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACT 2100
QY 2539 TTGGAAGAAATTCCTTCTTAAATCAAAAAACAAACAAAAACAAAAACAAAAACATTC 2598
Db 2101 TTGGAAGAAATTCCTTCTTAAATCAAAAAACAAACAAAAACAAAAACAAAAACATTC 2160

QY 2599 TAAATCTAGAGATAAATTTACTTTAAATTTCTTCAATTTTAGCAGTGATGATATGCAATAGT 2658
Db 2161 TAAATCTAGAGATAAATTTACTTTAAATTTCTTCAATTTTAGCAGTGATGATATGCAATAGT 2220
QY 2659 GCTGTAAAGGCTTTAACTGGGGAATATTTCCACCTGATTAATAGCCAGATTTCTACTGTAT 2718
Db 2221 GCTGTAAAGGCTTTAACTGGGGAATATTTCCACCTGATTAATAGCCAGATTTCTACTGTAT 2280
QY 2719 TCCAAAAGGCAATATTAAAGGTAGATAGATGATATTAGTATATATTGTTTACACATATTTT 2778
Db 2281 TCCAAAAGGCAATATTAAAGGTAGACAGATGATATTAGTATATATTGTTTACACATATTTT 2340
QY 2779 GGAATTAGAAGAACATACAGAAAGGAATTTAGGGGCTTAAACATTAAGCTGAAATGCATTTT 2838
Db 2341 GGAATTAGAAGAACATACAGAAAGGAATTTAGGGGCTTAAACATTAAGCTGAAATGCATTTT 2400
QY 2839 AGTATAAGGSCACAGTTGTATATTTTAAATCAATACCAATTTAAATTTTAAATTTT 2898
Db 2401 AGTATAAGGSCACAGTTGTATATTTTAAATCAATACCAATTTAAATTTTAAATTTT 2460
QY 2899 ACCTGTTAAAGAGATTTTACTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2958
Db 2461 ACCTGTTAAAGAGATTTTACTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2520
QY 2959 TATCAGGAATTTACTACTTTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3018
Db 2521 TATCAGGAATTTACTACTTTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2580
QY 3019 GAGGTATAATCAACAGAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGG 3078
Db 2581 GAGGTATAATCAACAGAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGG 2640
QY 3079 TGTCTAGAGAGATCTGTCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGT 3138
Db 2641 TGTCTAGAGAGATCTGTCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGT 2700
QY 3139 ACTCCAAATGATTTCTTTCTGCTGATATCTGCTGCTTCTCATATTTACTGAAAGCT 3198
Db 2701 ACTCCAAATGATTTCTTTCTGCTGATATCTGCTGCTTCTCATATTTACTGAAAGCT 2760
QY 3199 GCAATATTTAGTAAATACCTTCGGGATCACTGTCCTCCCATCTTCCTGTTAGAGCAAGT 3258
Db 2761 GCAATATTTAGTAAATACCTTCGGGATCACTGTCCTCCCATCTTCCTGTTAGAGCAAGT 2820
QY 3259 GAAGAGTTTAAAGAGGAGAGAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 3318
Db 2821 GAAGAGTTTAAAGAGGAGAGAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 2880
QY 3319 AACCTGTATTTCCCTTATGATGTCCTTTTGGAGACACTTAATTTTAAATACTTACTA 3378
Db 2881 AACCTGTATTTCCCTTATGATGTCCTTTTGGAGACACTTAATTTTAAATACTTACTA 2940
QY 3379 GCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAG 3438
Db 2941 GCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAG 3000
QY 3439 GCCTTTTCTTGGGATGATTTCTTCTTAAAGTTTGGGACATTAATAAAGTTAGTAC 3498
Db 3001 GCCTTTTCTTGGGATGATTTCTTCTTAAAGTTTGGGACATTAATAAAGTTAGTAC 3060
QY 3499 ATTTGTTGTACACAGTTGATTTCCAAATTTGATGATGGAGGAGAGGTGCTTAAGC 3558
Db 3061 ATTTGTTGTACACAGTTGATTTCCAAATTTGATGATGGAGGAGAGGTGCTTAAGC 3120
QY 3559 TGTAAGGCTTTCTTTGACTGCAATTTATAGAGATTTTACGTTTAAATTTTAAATTTTAAAGTGT 3618
Db 3121 TGTAAGGCTTTCTTTGACTGCAATTTATAGAGATTTTACGTTTAAATTTTAAATTTTAAAGTGT 3180
QY 3619 AAAACATCTGCTTTCTTCTTCTTACCTAGTCTGAAACATTTTAAATTTTAAATTTTAAAGTGT 3678
Db 3181 AAAACATCTGCTTTCTTCTTCTTACCTAGTCTGAAACATTTTAAATTTTAAATTTTAAAGTGT 3240
QY 3679 AATTAATTTTGAAGAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAGTGT 3705

|||||
3241 AATTAAATTTGAAAAAATAAAAAA 3267

Db 3241 AATTAAATTTGAAAAAATAAAAAA 3267

RESULT 5
AF198455 3550 bp mRNA linear PRI 10-JAN-2000
LOCUS
DEFINITION Homo sapiens epithelial protein lost in neoplasm alpha (EPLIN)
mRNA, complete cds.
ACCESSION AF198455
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3550)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE EPLIN, epithelial protein lost in neoplasm
JOURNAL Oncogene 18 (54): 7838-7841 (1999)
MEDLINE 20087188
PUBMED 10618726
REFERENCE 2 (bases 1 to 3550)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
FEATURES
source
1..3550
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
1..3550
/gene="EPLIN"
474..2276
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm alpha"
/protein_id="AA233756.1"
/db_xref="GI:6685008"
/translation="MENCIGESRHEVEKSEISNTASGKIEKYNVPLNRLKMFEG
EPQTQILRAQSSAGSRKISNSYSLDLEIGQLSSSTFSEKNRRNLELPRL
SETSIKIDRAQSSAGSRKISNSYSLDLEIGQLSSSTFSEKNRRNLELPRL
EGEKISANENSLAVRSTPAEDDSDPSQVQVHPKPLSPDPSRSLSESPPK
AMKFPQAPARETCVEQKTYPMERLLANQQVPHISCFPCSVCNNKLSLGTYSALHGR
IYCKPHNQIFKSGMIDEGFGRPHKDLWASKNEEILERPQALANARETPHSEGV
EDAPIAKGVLAASMEAKSSQEKEDKPAETKXLRIAWPPTELGSQSALEEGTKM
SKPKWPPEDDISKPEVDYDLDLKLRSSSLKRSRPTVAASQSTSVKSPKTVS
PPIRKGMSEQSEESVGGVAVRKOVENAKSKKNVGNKKTWONKESKGTGRSK
EGHSLMENLNVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTPAEFTTQNK
SQDELWEGVVKVLSVEEQIKENRYDEDEEE"
polyA_signal
3527..3532
/gene="EPLIN"
BASE COUNT 1153 a 716 c 773 g 908 t
ORIGIN
Query Match 83.68; Score 3098; DB 9; Length 3550;
Best Local Similarity 99.94; Pred. No. 0;
Matches 3388; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 302 AACACCCAAAATCTCTCCAGCACTTTAGAAAAGGGACCTGACTGTGTTAAAGAGAAG 361
Db 159 AACACCCAAAATCTCTCCAGCACTTTAGAAAAGGGACCTGACTGTGTTAAAGAGAAG 218
QY 362 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAACAGCAGCACT 421
Db 219 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAACAGCAGCACT 278
QY 422 GAGATTAGGCACAGACAGACCATCTCTCTGCTGAAGTGACAGCAACGCTGCTCTCGGA 481
Db 279 GAGATTAGGCACAGACAGACCATCTCTCTGCTGAAGTGACAGCAACGCTGCTCTCGGA 338

QY 482 GCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCTCCTCTGAA 541
Db 339 GCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCTCCTCTGAA 398
QY 542 GCCCTCGTTCAGGTCGATATCCCAACATCAAGGACGTCAGGATCTTAAAGACCACTCA 601
Db 399 GCCCTCGTTCAGGTCGATATCCCAACATCAAGGACGTCAGGATCTTAAAGACCACTCA 458
QY 602 ACAGAAAGTAAAAAAATGAAAAATTTGTCTAGGAGAAATCCAGGCATGAAAGTAAAAATCA 661
Db 459 ACAGAAAGTAAAAAAATGAAAAATTTGTCTAGGAGAAATCCAGGCATGAAAGTAAAAATCA 518
QY 662 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATTTCCGCTGAAC 721
Db 519 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATTTCCGCTGAAC 578
QY 722 AGGCTTAAGATGATGTTTGAAGAGGTGAACCACTCAAACTAAGATTTCTCGGCGCCAA 781
Db 579 AGGCTTAAGATGATGTTTGAAGAGGTGAACCACTCAAACTAAGATTTCTCGGCGCCAA 638
QY 782 AGCCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGAA 841
Db 639 AGCCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGAA 698
QY 842 ATAGGCCAGGTCAAGTCTCTGATCTTTTACATTTGACTCGGAGAAAAATAGAGTAGACGA 901
Db 699 ATAGGCCAGGTCAAGTCTCTGATCTTTTACATTTGACTCGGAGAAAAATAGAGTAGACGA 758
QY 902 AATCTGGAACCTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAG 961
Db 759 AATCTGGAACCTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAG 818
QY 962 GCAGCTGTGTCACCAACAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 1021
Db 819 GCAGCTGTGTCACCAACAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 878
QY 1022 GCGCAATCAAAATTCATAAAATGGAGCAAAAGGAGATGTGCCCCAGGTCTCTCAGGTC 1081
Db 879 GCGCAATCAAAATTCATAAAATGGAGCAAAAGGAGATGTGCCCCAGGTCTCTCAGGTC 938
QY 1082 TGCAATCAACCCATCAGAAAGGGAAAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCGT 1141
Db 939 TGCAATCAACCCATCAGAAAGGGAAAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCGT 998
QY 1142 TCCACCCCTGCGGAAGATGACTCC --GTGACTCCAGGTTAAGAGTGAGGTTCAACAG 1198
Db 999 TCCACCCCTGCGGAAGATGACTCCAGGTTAAGAGTGAGGTTCAACAG 1058
QY 1199 CCTGTCCATCCCAAGCCCACTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCT 1258
Db 1059 CCTGTCCATCCCAAGCCCACTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCT 1118
QY 1259 CCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGAGCTGCGTGAATGTCTCAG 1318
Db 1119 CCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGAGCTGCGTGAATGTCTCAG 1178
QY 1319 AAGACAGTCTATCCAAATGGAGCGTCTTTTGCCCAACCCAGCAGGTGTTTTCACATCAGTGC 1378
Db 1179 AAGACAGTCTATCCAAATGGAGCGTCTTTTGCCCAACCCAGCAGGTGTTTTCACATCAGTGC 1238
QY 1379 TTCGTTGCTCTTATTGCAACCAAACTCAGTCTAGGAAACATATGCATCTTTTACATGGA 1438
Db 1239 TTCGTTGCTCTTATTGCAACCAAACTCAGTCTAGGAAACATATGCATCTTTTACATGGA 1298
QY 1439 AGAATCTATTGTAAGCTCAGTCTCAATCAACTCTTTTAAATCTAAGGGCACTATGATGAA 1498
Db 1299 AGAATCTATTGTAAGCTCAGTCTCAATCAACTCTTTTAAATCTAAGGGCACTATGATGAA 1358
QY 1499 GCGTTTGGGCACAGACCAACAGGATCTTATGGCAAGCAAAATGAAACGCAAGAGATT 1558
Db 1359 GCGTTTGGGCACAGACCAACAGGATCTTATGGCAAGCAAAATGAAACGCAAGAGATT 1418

Qy	1559	TTGGAGAGACCAGCCAGCTTGC	AAATGCAAGGGAGAGCCCTCACAGCC	CAGGGGTAGAA	1619		
Db	1419	TTGGAGAGACCAGCCAGCTTGC	AAATGCAAGGGAGAGCCCTCACAGCC	CAGGGGTAGAA	1478		
Qy	1619	GATGCCCTATTGCTAAGGTGG	TGTCCTGCTGCAAGTATGGAAG	CCAGGCTCTCTCT	1678		
Db	1479	GATGCCCTATTGCTAAGGTGG	TGTCCTGCTGCAAGTATGGAAG	CCAGGCTCTCTCT	1538		
Qy	1679	CAGCAGGAGAAAGGAGACAAG	CCAGCTGAAACCAAGAAAGCTG	AGGATCGCTTGCCACACC	1738		
Db	1539	CAGCAGGAGAAAGGAGACAAG	CCAGCTGAAACCAAGAAAGCTG	AGGATCGCTTGCCACACC	1598		
Qy	1739	CCCACTGAACCTTGAAGATT	CAGGAAGTGCCTTGGAGGAAG	GGATCAAAATGTCAAAAGCCC	1798		
Db	1599	CCCACTGAACCTTGAAGATT	CAGGAAGTGCCTTGGAGGAAG	GGATCAAAATGTCAAAAGCCC	1658		
Qy	1799	AAATGGCTCTCTGAAGACG	AAATCAGCAAGCCCAAGTTC	CTGAGGATGTTCGATCTAGAT	1858		
Db	1659	AAATGGCTCTCTGAAGACG	AAATCAGCAAGCCCAAGTTC	CTGAGGATGTTCGATCTAGAT	1718		
Qy	1859	CTGAAGAAAGCTAAGACGAT	CTCTTCACTGAAGGAAAGAA	CGCCCAATTCAGCTGTAGCA	1918		
Db	1719	CTGAAGAAAGCTAAGACGAT	CTCTTCACTGAAGGAAAGAA	CGCCCAATTCAGCTGTAGCA	1778		
Qy	1919	GCTTCAATTTCAAAGACCT	CTGTCAAGAGCCCAAAA	CTGTGCCCCA	CCCTATCAGGAAA	1978	
Db	1779	GCTTCAATTTCAAAGACCT	CTGTCAAGAGCCCAAAA	CTGTGCCCCA	CCCTATCAGGAAA	1838	
Qy	1979	GGCTGGAGCATGTCCAGAG	CAGAGTGAAGTCTGTGGTG	GGAGAGTTGCAGAAAGGAAA	2038		
Db	1839	GGCTGGAGCATGTCCAGAG	CAGAGTGAAGTCTGTGGTG	GGAGAGTTGCAGAAAGGAAA	1898		
Qy	2039	CAAGTGGAAAAATGCCAAG	GGCTTCTTAAGAAAGAA	TGGGAATGGGAA	AAAAACAACTCGCAA	2098	
Db	1899	CAAGTGGAAAAATGCCAAG	GGCTTCTTAAGAAAGAA	TGGGAATGGGAA	AAAAACAACTCGCAA	1958	
Qy	2099	AACAAAGAACTTAAGGAG	AGAGAGAGGAAGTAAGAA	GGTCTATAGTTGGGAGATG	2158		
Db	1959	AACAAAGAACTTAAGGAG	AGAGAGAGGAAGTAAGAA	GGTCTATAGTTGGGAGATG	2018		
Qy	2159	GAGATGAGAACTTGTATG	AAAAATGGTGAGACTCCG	ATCAAGATGATACACACTTCTC	2218		
Db	2019	GAGATGAGAACTTGTATG	AAAAATGGTGAGACTCCG	ATCAAGATGATACACACTTCTC	2078		
Qy	2219	AAACAACAATCTCCA	GAAGACCCCAAGTCTCT	GAAATTTGGT	CGAGTTTGTAGACA	CAACC	2278
Db	2079	AAACAACAATCTCCA	GAAGACCCCAAGTCTCT	GAAATTTGGT	CGAGTTTGTAGACA	CAACC	2138
Qy	2279	TTTGCTGAAGAAATTCAT	CTACTAGAAATCGA	AAATCCAGGATCGGAACCTCTGGGAGGA	2338		
Db	2139	TTTGCTGAAGAAATTCAT	CTACTAGAAATCGA	AAATCCAGGATCGGAACCTCTGGGAGGA	2198		
Qy	2339	GAAGTGGTCAAGAGCTCT	CTGTGGAGAGACAGAT	TAAGAGAAATCGGTATTATGATGAG	2398		
Db	2199	GAAGTGGTCAAGAGCTCT	CTGTGGAGAGACAGAT	TAAGAGAAATCGGTATTATGATGAG	2258		
Qy	2399	GATGAGGATGAAGAGTGA	CAAAATGCAATGCTGGCCCT	TAAATTCATGTTAGTTGTTA	2458		
Db	2259	GATGAGGATGAAGAGTGA	CAAAATGCAATGCTGGCCCT	TAAATTCATGTTAGTTGTTA	2318		
Qy	2459	CGGAGCCACTGCCCCCT	TTTGTCAAAATGTGAT	GCACATAAGCAGGAT	TCCGAGCATGAAATG	2518	
Db	2319	CGGAGCCACTGCCCCCT	TTTGTCAAAATGTGAT	GCACATAAGCAGGAT	TCCGAGCATGAAATG	2378	
Qy	2519	TAATTTACTTTGGAGTAA	CTTTTGGAAAGAAATTC	CTTCTTAAATCAAAA	CAAAACAAA	2578	
Db	2379	TAATTTACTTTGGAGTAA	CTTTTGGAAAGAAATTC	CTTCTTAAATCAAAA	CAAAACAAA	2438	
Qy	2579	AAAAACA	CAAAAAACACATCTCT	TAATACTAGAGATA	CTTTTACTTAAATTTCTCATTTTAG	2638	
Db	2439	AAAAACA	CAAAAAACACATCTCT	TAATACTAGAGATA	CTTTTACTTAAATTTCTCATTTTAG	2498	
Qy	2639	CAGTGATGATATG	CATAAGTGTGTAAG	CGCTTGTAACTGGG	AAAAATATTTCCACCTG	ATAA	2698

AX017487
LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 31 from Patent WO9947655.
ACCESSION AX017487
VERSION AX017487.1 GI:10042284
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pilarczyk, C.
TITLE Human nucleic acid sequences from normal breast tissue
JOURNAL Patent: WO 9947655-A 31 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..2783
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 901 a 526 c 609 g 747 t
ORIGIN
Query Match 71.4%; Score 2647; DB 6; Length 2783;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2697; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAATGGACAAAGAGAAAT 1060
DB 40 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAATGGACAAAGAGAAAT 99
QY 1061 GTGCCCCAGGTCCTGAGTCTGCATCACCATCAGGAGGGGAAAGATTTCTGCAAT 1120
DB 100 GTGCCCCAGGTCCTGAGTCTGCATCACCATCAGGAGGGGAAAGATTTCTGCAAT 159
QY 1121 GAGAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 1180
DB 160 GAGAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 219
QY 1181 AAGAGTGAAGTTCAACAGCTTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 1240
DB 220 AAGAGTGAAGTTCAACAGCTTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 279
QY 1241 AGCTTTCTGAAGTCTCTCCCAAGCAATGAAGATTTTCAGGACCTTGCAGAGAG 1300
DB 280 AGTCTTTCTGAAGTCTCTCCCAAGCAATGAAGATTTTCAGGACCTTGCAGAGAG 339
QY 1301 ACCTCGTGGAAATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCGAG 1360
DB 340 ACCTCGTGGAAATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCGAG 399
QY 1361 GTGTTTCACATCAGTGTGTCGTTGCTCCTATTCGCAACAACTCAGTCTAGGAACA 1420
DB 400 GTGTTTCACATCAGTGTGTCGTTGCTCCTATTCGCAACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAGATCTATTGTAAGCCTCATTCAATCAACTCTTTAAATCT 1480
DB 460 TATGCATCTTTACATGGAGATCTATTGTAAGCCTCATTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCGAGCAAA 1540
DB 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCGAGCAAA 579
QY 1541 AATGAAACGAGAGATTTGGAGAGACCCAGCTTGCAAATGCAAGGGAGACCCCT 1600
DB 580 AATGAAACGAGAGATTTGGAGAGACCCAGCTTGCAAATGCAAGGGAGACCCCT 639
QY 1601 CACAGCCAGGGGTAGAGATGCCCTTATGCTAAGGTGGGTGCTCCTGGCTGCAAGTATG 1660
DB 640 CACAGCCAGGGGTAGAGATGCCCTTATGCTAAGGTGGGTGCTCCTGGCTGCAAGTATG 699

QY 1661 GAAGCCCAAGCCTCTCTCTCAGCAGGAGAGAAAGCAAGCCAGCTGAAACCAAGAGCTG 1720
DB 700 GAAGCCCAAGCCTCTCTCTCAGCAGGAGAGAAAGCAAGCCAGCTGAAACCAAGAGCTG 759
QY 1721 AGATTCGCTTGGCCACCCCTCCTGAAAGTTTCAGGAAGTGCCTTTGGAGGAAGGG 1780
DB 760 AGATTCGCTTGGCCACCCCTCCTGAAAGTTTCAGGAAGTGCCTTTGGAGGAAGGG 819
QY 1781 ATCAAAATGTCAAGGCCAAATGSCCTCCTGAAGAGCGAAATCAGCAAGCCCGAAGTTCT 1840
DB 820 ATCAAAATGTCAAGGCCAAATGSCCTCCTGAAGAGCGAAATCAGCAAGCCCGAAGTTCT 879
QY 1841 GAGGATCTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 1900
DB 880 GAGGATCTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 939
QY 1901 CGCCCATCTAGTGTAGCAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAAATCTGTG 1960
DB 940 CGCCCATCTAGTGTAGCAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAAATCTGTG 999
QY 1961 TCCCCACCTATCAGGAAGGCTCGACATGTCTAGACAGAGTCAAGAGTCTGTGGGTGGA 2020
DB 1000 TCCCCACCTATCAGGAAGGCTCGACATGTCTAGACAGAGTCAAGAGTCTGTGGGTGGA 1059
QY 2021 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 2080
DB 1060 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 1119
QY 2081 GGAAAAACAACTGGGCAAAACAAAGAAATCTTAAAGGAGACAGCGGAGAGAGAAAGGAA 2140
DB 1120 GGAAAAACAACTGGGCAAAACAAAGAAATCTTAAAGGAGACAGCGGAGAGAGAAAGGAA 1179
QY 2141 GGTCTATGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAATGGTGCAAGTCCGATGAA 2200
DB 1180 GGTCTATGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAATGGTGCAAGTCCGATGAA 1239
QY 2201 GATGATAACAGCTTCTCBAACCAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTGC 2260
DB 1240 GATGATAACAGCTTCTCBAACCAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTGC 1299
QY 2261 AGTTTTGTAGAACACACCTTTGCTGAAGAAATTCACACTCAGAAATCAGAAATCCCAGGAT 2320
DB 1300 AGTTTTGTAGAACACACCTTTGCTGAAGAAATTCACACTCAGAAATCAGAAATCCCAGGAT 1359
QY 2321 GTGGAACCTGTGGAGGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2380
DB 1360 GTGGAACCTGTGGAGGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 1419
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGTGAACAATTCGAATGATGCTGGGCCCTT 2440
DB 1420 AATCGGTATTATGATGAGGATGAGGATGAAGTGAACAATTCGAATGATGCTGGGCCCTT 1479
QY 2441 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAG 2500
DB 1480 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAG 1539
QY 2501 GTATCCAGCATGAAATGTAAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTAA 2560
DB 1540 GTATCCAGCATGAAATGTAAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTAA 1599
QY 2561 AATCAAAACAAACAAAAAACAACAAAAACACATTTCTAAATCTAGAGATAAATTTAC 2620
DB 1600 AATCAAAACAAACAAAAAACAACAAAAACACATTTCTAAATCTAGAGATAAATTTAC 1659
QY 2621 TTAATTTCTTCTTTTAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2680
DB 1660 TTAATTTCTTCTTTTAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1719
QY 2681 AAATATTCACCTGTATATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGT 2740
DB 1720 AAATATTCACCTGTATATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGT 1779

Db 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAA 579
Qy 1541 AATGAAAACGAAGAGATTTTGGAGAGACCAAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 1600
Db 580 AATGAAAACGAAGAGATTTTGGAGAGACCAAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 639
Qy 1601 CACAGCCAGGGGTAGAGATGCCCTTATTGCTAAGGTGGGTGCTCCTGGCTGCAAGTATG 1660
Db 640 CACAGCCAGGGGTAGAGATGCCCTTATTGCTAAGGGGGGTGCTCCTGGCTGCAAGTATG 699
Qy 1661 GAAGCCAAAGCCCTCTCTCAGCAGAGAGAGAAAGACAAGCCAGCTGAAACCAAGAAAGCTG 1720
Db 700 GAAGCCAAAGCCCTCTCTCAGCAGAGAGAGAGAAAGACAAGCCAGCTGAAACCAAGAAAGCTG 759
Qy 1721 AGSATCGCTGGCCACCCCACTGAACCTTGGAAGTTCAGAAAGTGCCTTGGAGGAAGGG 1780
Db 760 AGSATCGCTGGCCACCCCACTGAACCTTGGAAGTTCAGAAAGTGCCTTGGAGGAAGGG 819
Qy 1781 ATCAAAATGTCAAAAGCCCAATGCCCTCTGAGAGAGAAATCAGCAAGCCCGAAGTTCCT 1840
Db 820 ATCAAAATGTCAAAAGCCCAATGCCCTCTGAGAGAGAAATCAGCAAGCCCGAAGTTCCT 879
Qy 1841 GAGGATCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAGAAAGC 1900
Db 880 GAGGATCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAGAAAGC 939
Qy 1901 CGCCCATCTAGTGAAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAATCTGTG 1960
Db 940 CGCCCATCTAGTGAAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAATCTGTG 999
Qy 1961 TCCCACTCTATCAGGAAGGCTGGAGCATGTCAAGACAGAGTGAAGAGTCTGTGGTGGA 2020
Db 1000 TCCCACTCTATCAGGAAGGCTGGAGCATGTCAAGACAGAGTGAAGAGTCTGTGGTGGA 1059
Qy 2021 AGAGTTGAGAAAGAAACAAAGTGGAAATTCGCAAGCTTCTTAAGAGAAATGGAATGTG 2080
Db 1060 AGAGTTGAGAAAGAAACAAAGTGGAAATTCGCAAGCTTCTTAAGAGAAATGGAATGTG 1119
Qy 2081 GGAAAAACAACCTGGCAAAACAAAGAACTTAAAGGAGAGACAGGGAAGAGAAAGGAA 2140
Db 1120 GGAAAAACAACCTGGCAAAACAAAGAACTTAAAGGAGAGACAGGGAAGAGAAAGGAA 1179
Qy 2141 GGTATAGTTTGGAGATGGAGATGAGAACTTGTAGAAATGGTGCAAGCTCCGATGAA 2200
Db 1180 GGTATAGTTTGGAGATGGAGATGAGAACTTGTAGAAATGGTGCAAGCTCCGATGAA 1239
Qy 2201 GATGATAACAGCTTCTCACAACAACTCTCACAAGACCCCAAGTCTCTGAATGGTGG 2260
Db 1240 GATGATAACAGCTTCTCACAACAACTCTCACAAGACCCCAAGTCTCTGAATGGTGG 1299
Qy 2261 AGTTTGTAGACAACACCTTTGTCTGAAGAAATCACTACTCAGAAATCAGAAATCCCAAGAT 2320
Db 1300 AGTTTGTAGACAACACCTTTGTCTGAAGAAATCACTACTCAGAAATCAGAAATCCCAAGAT 1359
Qy 2321 GTGGAATCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGA 2380
Db 1360 GTGGAATCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGA 1419
Qy 2381 AATCGGTATTATGATGAGGATGAGGATGAGAGTGCATAATTCGAATGATGCTGGGCCCTT 2440
Db 1420 AATCGGTATTATGATGAGGATGAGGATGAGAGTGCATAATTCGAATGATGCTGGGCCCTT 1479
Qy 2441 AAATTCATGTTAGTGTAGGAGCCACTGCCCCCTTGTCAAAATGTGATGCACATAAGCAG 2500
Db 1480 AAATTCATGTTAGTGTAGGAGCCACTGCCCCCTTGTCAAAATGTGATGCACATAAGCAG 1539
Qy 2501 GTATCCAGCATGAATGTAAATTAATCTGGAAGTAACTTTGGAAAAAGAAATCTCTCTTAA 2560
Db 1540 GTATCCAGCATGAATGTAAATTAATCTGGAAGTAACTTTGGAAAAAGAAATCTCTCTTAA 1599
Qy 2561 AATCAAAACAAACAAACAAACAAACAAACAAACAACTTAAATACTAGAGATACTTTAC 2620
Db 1600 AATCAAAACAAACAAACAAACAAACAAACAAACAACTTAAATACTAGAGATACTTTAC 1659

Qy 2621 TTAAATTTCTTCAATTTTAGCAGTGATGATATGATCAATAGTGTGTAAAGCTTTGTAACCTGGG 2680
Db 1660 TTAAATTTCTTCAATTTTAGCAGTGATGATATGATCAATAGTGTGTAAAGCTTTGTAACCTGGG 1719
Qy 2681 AATATATCCACCTGATATAGCCAGATTTCTACTGTATTTCCCAAAAGCAATATTAAGGT 2740
Db 1720 AATATATCCACCTGATATAGCCAGATTTCTACTGTATTTCCCAAAAGCAATATTAAGGT 1779
Qy 2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAAATACAGAAAG 2800
Db 1780 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAAATACAGAAAG 1839
Qy 2801 GAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCAGTTTGTGTA 2860
Db 1840 GAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCAGTTTGTGTA 1899
Qy 2861 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTAAGAGATTTATTAGT 2920
Db 1900 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTAAGAGATTTATTAGT 1959
Qy 2921 CTTTAAATTTTGTAGTTAAATTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2980
Db 1960 CTTTAAATTTTGTAGTTAAATTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2019
Qy 2981 GTCTGTCTCTAAACTACATCTGAACCTGACGCTCTGAGGTATAATACACAGAGACAC 3040
Db 2020 GTCTGTCTCTAAACTACATCTGAACCTGACGCTCTGAGGTATAATACACAGAGACAC 2079
Qy 3041 TTTTGTAGGCAATTTGAAAAACCAACCTACCTCTCTGCTGTGTAGAGAGATCTGCTGTCT 3100
Db 2080 TTTTGTAGGCAATTTGAAAAACCAACCTACCTCTCTGCTGTGTAGAGAGATCTGCTGTCT 2139
Qy 3101 CCCAAATAGCTTTGTATCTGCCAGTCAATTTACTGTACTTCCAAATGATTCCTTTCTTT 3160
Db 2140 CCCAAATAGCTTTGTATCTGCCAGTCAATTTACTGTACTTCCAAATGATTCCTTTCTTT 2199
Qy 3161 TCTGTGTATCTGCTGTCTCTCATAAATTAAGTGAAGTGTCAATATTTTAGTAATACCTTC 3220
Db 2200 TCTGTGTATCTGCTGTCTCTCATAAATTAAGTGTCAATATTTTAGTAATACCTTC 2259
Qy 3221 GGGATCACTGTCCCCCATCTTCCGTGTGTAGCAAAAGTGAAGAGTTTAAAGGAGAGAA 3280
Db 2260 GGGATCACTGTCCCCCATCTTCCGTGTGTAGCAAAAGTGAAGAGTTTAAAGGAGAGAA 2319
Qy 3281 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGAT 3340
Db 2320 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGAT 2379
Qy 3341 GTCCCCCTTTTGAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATGATTTTT 3400
Db 2380 GTCCCCCTTTTGAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATGATTTTT 2439
Qy 3401 ATCAGATATTTCTCAGGCTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTT 3460
Db 2440 ATCAGATATTTCTCAGGCTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTT 2499
Qy 3461 CTAGCTTTAAGGTTGGGACATTTATAAATCTGTAGTACATTTGTGTACACAGTTGATAT 3520
Db 2500 CTAGCTTTAAGGTTGGGACATTTATAAATCTGTAGTACATTTGTGTGTACACAGTTGATAT 2559
Qy 3521 TCCAAATGTATGATGGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGTC 3580
Db 2560 TCCAAATGTATGATGGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGTC 2619
Qy 3581 ATTTATAGAGATTTAGCTTTTAAATTTTGTAGAGATGTAAACCACTTCTGCTTCTTAGTC 3640
Db 2620 ATTTATAGAGATTTAGCTTTTAAATTTTGTAGAGATGTAAACCACTTCTGCTTCTTAGTC 2679
Qy 3641 TTACTAGTCTGAAACATTTTATTTCAATAAGATTTTAAATTTAAATTTGAAAAAAA 3698
Db 2680 TTACTAGTCTGAAACATTTTATTTCAATAAGATTTTAAATTTAAATTTGAAAAAAA 2737


```

RESULT 8
AK000372
LOCUS Homo sapiens cDNA FJ20365 f1s, clone HEP17877.
DEFINITION Homo sapiens cDNA FJ20365 f1s, clone HEP17877.
ACCESSION AK000372
VERSION AK000372.1 GI:7020415
KEYWORDS oligo cloning; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,
Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2536)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology; Shirohane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail: conai@ims.u-tokyo.ac.jp,
Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5' & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. 2536
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP17877"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/notes="cloning vector pME18SFL3"
85..2199
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91120.1"
/db_xref="GI:7020416"
/translation="MESSPFRNRQWTSLSRVTAKELSLVNKNKSSAIVEIRSKYOKA
AETNMKRSNTENISQHFRTGLTVLKKWENPGLGAESHTDSLNSSTIRHEAD
HPAEVTSAAAGAKADQEQIHPRLSRSPPEALVQGRYPHFKDGEDLKDSTESKK
MENCLGSRHEVEKSEISSENTDASGKIEKYNVPLNRLKMFKEGPTQILRAQSRK
ASGRKISENSYSLDDLEIGRQGLSSSTFDSEKVESRRNLJELPLRSETSIKDRMAKYQA
AVKQSSSTNYTNELKASGGEIKHWQENYVPGPEVCIHQEGEKISANENSLAV
RSTPAEDSDSQVKEVQVHPKPLSPDRASSLSSESPPKAMKFOAPARETCVE
CQKTVYPMERLLANQOVFHISCFRCVCNNKLSLGTYSALHGRFYCKPHNOLFASKG
NYDEGFGRPHKDLWASKNENELIRPAQLANARETSPGVEDPIAKVGLAASG
EAKASQOEKEDPASTKLRIRAMPPTTELGGSSALEEGIKMSKPKWPFDEISKPE
VPEDVDLLKRRSLRSLRAMPPTTELGGSSALEEGIKMSKPKWPFDEISKPE
SVGRVAERKQKAKSKNGKNGVKTITWQNKESKGTGKRKEGHSLEWENLIVEN
GADSDDDNSFLKQHLKPNLSL"
BASE COUNT 863 a 551 c 612 g 510 t
ORIGIN
Query Match 64.4%; Score 2386; DB 9; Length 2536;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2506; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 63 GACCTGTGACAGGCTGGTAGACCGCAGAGGAAGCGCGCTTTTAGCCAGGTATTTCAG 122
DB 11 GACCTGTGACAGGCTGGTAGACCGCAGAGGAAGCGCGCTTTTAGCCAGGTATTTCAG 70

```

```

QY 123 TGTCTGTAGACAGATGGAATCATCTCCATTTAATAGACGGCAATGGACCTCATTATCAT 182
DB 71 TGTCTGTAGACAGATGGAATCATCTCCATTTAATAGACGGCAATGGACCTCATTATCAT 130
QY 183 TGAGCGGTAAACAGGCAAGAACTTTCTTGTGTCAACAAGAACAGTCTATCGGCTATTGTGG 242
DB 131 TGAGCGGTAAACAGGCAAGAACTTTCTTGTGTCAACAAGAACAGTCTATCGGCTATTGTGG 190
QY 243 AAATATTCTCCAGTACCAGAAAGCAGCTGAAGAAAACAAACATGGAGAAAGAGAAGTA 302
DB 191 AAATATTCTCCAGTACCAGAAAGCAGCTGAAGAAAACAAACATGGAGAAAGAGAAGTA 250
QY 303 ACACCGAAATCTCTCCAGCACTTTAGAAAAGGGACCTGTGCTGTGTTTAAAGAAAGAGT 362
DB 251 ACACCGAAATCTCTCCAGCACTTTAGAAAAGGGACCTGTGCTGTGTTTAAAGAAAGAGT 310
QY 363 GGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGACAGCAGCACTG 422
DB 311 GGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGACAGCAGCACTG 370
QY 423 AGATTAGGCACAGAGCAGACCATCTCTCTGTGTAAGTGAACAGCCACGCTGCTTCTGGAG 482
DB 371 AGATTAGGCACAGAGCAGACCATCTCTCTGTGTAAGTGAACAGCCACGCTGCTTCTGGAG 430
QY 483 CCAAAGCTGACCAAGAAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAG 542
DB 431 CCAAAGCTGACCAAGAAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAG 490
QY 543 CCTCTGTTGAGGTGCTATTCCTCCACATCAAGGACGGTGAAGATCTTAAAGACCACTCAA 602
DB 491 CCTCTGTTGAGGTGCTATTCCTCCACATCAAGGACGGTGAAGATCTTAAAGACCACTCAA 550
QY 603 CAGAAAGTAAATAATGGAAATTTGCTAGGAGAAATCCAGGATGAGTAGAAAAATCAG 662
DB 551 CAGAAAGTAAATAATGGAAATTTGCTAGGAGAAATCCAGGATGAGTAGAAAAATCAG 610
QY 663 AAATCAGTGAAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGACCA 722
DB 611 AAATCAGTGAAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGACCA 670
QY 723 GCGTTAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTTCTCCGGGCCCAA 782
DB 671 GCGTTAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTTCTCCGGGCCCAA 730
QY 783 GCGAAGTCCAAAGTGAAGAAAGATCTCTGAAAAACAGTATTTCTTAGATGACCTGGAAA 842
DB 731 GCGAAGTCCAAAGTGAAGAAAGATCTCTGAAAAACAGTATTTCTTAGATGACCTGGAAA 790
QY 843 TAGGCCCAGGTGAGTGTCTCTTCTAGATTGACTCGGAGAAAAATGAGAGTAGACGAA 902
DB 791 TAGGCCCAGGTGAGTGTCTCTTCTAGATTGACTCGGAGAAAAATGAGAGTAGACGAA 850
QY 903 ATCTGGAACCTTCCACGCTCTCAGAAAACTCTATAAAGATCGAATGCGCAAGTACCAGG 962
DB 851 ATCTGGAACCTTCCACGCTCTCAGAAAACTCTATAAAGATCGAATGCGCAAGTACCAGG 910
QY 963 CAGCTGTGTCCAAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGCTG 1022
DB 911 CAGCTGTGTCCAAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGCTG 970
QY 1023 GCGAAATCAAAATTCATAAATGAGCAAAAGAGAAATGTGCCCCAGGTCTGAGGCT 1082
DB 971 GCGAAATCAAAATTCATAAATGAGCAAAAGAGAAATGTGCCCCAGGTCTGAGGCT 1030
QY 1083 GCATCAACCCATCAGAGGGGAAAGATTTCTGCAATAGAAATAGCCTGCGCAGTCCGTT 1142
DB 1031 GCATCAACCCATCAGAGGGGAAAGATTTCTGCAATAGAAATAGCCTGCGCAGTCCGTT 1090
QY 1143 CCACCCCTCGCGAAGATGACTCCCGTGTACTCCAGGTCTAAGAGTGAAGTTCAACAGCCTG 1202
DB 1091 CCACCCCTCGCGAAGATGACTCCCGTGTACTCCAGGTCTAAGAGTGAAGTTCAACAGCCTG 1150
QY 1203 TCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCTCCTC 1262

```

Db 1151 TCCATCCAGCCACTAAGTCAGATTCAGAGCTCCAGTCTTCTGAAAGTTCTCCTC 1210
QY 1263 CCAAGCAATGAAGAGTTTTCAGGACCTGCAAGAGACCTGCGTGAATGTGAGAGA 1322
Db 1211 CCAAGCAATGAAGAGTTTTCAGGACCTGCAAGAGACCTGCGTGAATGTGAGAGA 1270
QY 1323 CAGTCTATCCATGAGCGTCTCTTGGCCACACAGCAGCGTGTTCACATCAGCTGCTTCC 1382
Db 1271 CAGTCTATCCATGAGCGTCTCTTGGCCACACAGCAGCGTGTTCACATCAGCTGCTTCC 1330
QY 1383 GTTGCTCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTATCATGGAAGAA 1442
Db 1331 GTTGCTCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTATCATGGAAGAA 1390
QY 1443 TCTATTGAAGCCTCAGTCTCAATCAACTCTTTTAAATCTTAAGGGCACTATGATGAAGGCT 1502
Db 1391 TCTATTGAAGCCTCAGTCTCAATCAACTCTTTTAAATCTTAAGGGCACTATGATGAAGGCT 1450
QY 1503 TTGGCCACAGACCAACCAAGATCTATGGCCAGCAAAATGAAACCGAAGAGATTTGG 1562
Db 1451 TTGGCCACAGACCAACCAAGATCTATGGCCAGCAAAATGAAACCGAAGAGATTTGG 1510
QY 1563 AGAGACCAAGCCAGCTTGCATATGCAAGGAGACCCCTCAGAGCCAGGGGTAGAAGATG 1622
Db 1511 AGAGACCAAGCCAGCTTGCATATGCAAGGAGACCCCTCAGAGCCAGGGGTAGAAGATG 1570
QY 1623 CCCCTATTGCTAAGTGGGTGCTTGGTGCAGTATGGAAGCCAGCCCTCTCTCAGC 1682
Db 1571 CCCCTATTGCTAAGTGGGTGCTTGGTGCAGTATGGAAGCCAGCCCTCTCTCAGC 1630
QY 1683 AGGAGAGGAGACCAAGCAGCTGAAACCAAGAGCTGAGATGCGTGGCCACCCGCCA 1742
Db 1631 AGGAGAGGAGACCAAGCAGCTGAAACCAAGAGCTGAGATGCGTGGCCACCCGCCA 1690
QY 1743 CTGAACCTTGGAGTTTCAGAGAGTCTTGGAGAGGATCAAAATGTCAGAGCCCAAT 1802
Db 1691 CTGAACCTTGGAGTTTCAGAGAGTCTTGGAGAGGATCAAAATGTCAGAGCCCAAT 1750
QY 1803 GGCCTCTGAGACGAAATCAGAACGCCGAGTCTCAGGATGTCATCTAGATCTGA 1862
Db 1751 GGCCTCTGAGACGAAATCAGAACGCCGAGTCTCAGGATGTCATCTAGATCTGA 1810
QY 1863 AGAAGCTAAGAGCTTCTTCTCAGTGAAGGAAAGCGGCCATTCCTCTAGCAGCTT 1922
Db 1811 AGAAGCTAAGAGCTTCTTCTCAGTGAAGGAAAGCGGCCATTCCTCTAGCAGCTT 1870
QY 1923 CATTTCAAGACCTCTGTCAGAGCCCAAAACTGTGTCGCCACCTATCAGGAAAGGCT 1982
Db 1871 CATTTCAAGACCTCTGTCAGAGCCCAAAACTGTGTCGCCACCTATCAGGAAAGGCT 1930
QY 1983 GGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAGAGTTCAGAAAAGGAAACAAG 2042
Db 1931 GGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAGAGTTCAGAAAAGGAAACAAG 1990
QY 2043 TGGAAATGCCAAGCTTCTTAAGAGAAATGGAATGTGGGAAAAACAACCTGGCAAAACA 2102
Db 1991 TGGAAATGCCAAGCTTCTTAAGAGAAATGGAATGTGGGAAAAACAACCTGGCAAAACA 2050
QY 2103 AGAATCTTAAGAGAGACAGGAGAGAGAGTAAGAGAGTCAATAGTTTGAGATGAGA 2162
Db 2051 AGAATCTTAAGAGAGACAGGAGAGAGAGTAAGAGAGTCAATAGTTTGAGATGAGA 2110
QY 2163 ATGAGAATCTTGTAGAAATGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAAC 2222
Db 2111 ATGAGAATCTTGTAGAAATGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAAC 2170
QY 2223 AACAACTCTCCAGAACCCAGTCTCTGAATGTGTCAGTTTGTAGACAAACCTTTTG 2282
Db 2171 AAC-ATCTCCAGAACCCAGTCTCTGAATGTGTCAGTTTGTAGACAAACCTTTTG 2229
QY 2283 CTGAAGAATTTACTACTCAGATCAGAAATCCAGGATGTGGAATCTGCGAGGAGAG 2342

2230 CTGAAGATTCTACTACTCAGATCCAGAAATCCAGGATCTGGAACCTCTGGAGGAGAG 2289
2343 TGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGATG 2402
2290 TGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGATG 2349
2403 AGATGAGAGAGTACAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTTAGCGA 2462
2350 AGATGAGAGAGTACAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTTAGCGA 2409
2463 GCCACTGCCCTTTGTCAAATGATGATGCAATGATGCAATGATGCAATGATGCAATGATG 2522
2410 GCCACTGCCCTTTGTCAAATGATGATGCAATGATGCAATGATGCAATGATGCAATGATG 2469
2523 TTACTTGGAGTAACTTTTGGAAAGAAATTCCTTCTTAAATCAAAAA 2569
2470 TTACTTGGAGTAACTTTTGGAAAGAAATTCCTTCTTAAATCAAAAA 2516

RESULT 9
AF157325
LOCUS Homo sapiens sterol regulatory element binding protein 3 (SREBP3)
DEFINITION mRNA, complete cds.
ACCESSION AF157325
VERSION AF157325.1 GI:7688700
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
AUTHORS Huang, Q., Gu, J., Ren, S., Jin, W., Gu, Y., Gu, W., Dong, H., Yu, Y., Fu, G., Wang, Y., Chen, Z. and Han, Z.
TITLE A novel gene expressed in the human hypothalamus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3348)
AUTHORS Ren, S., Shi, J., Huang, C., Jiang, C., Li, Y., Zhou, J., Yu, Y., Xu, S., Wang, Y., Fu, G., Chen, Z. and Han, Z.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1999) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
FEATURES
source
1..3348
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
1..3348
/gene="SREBP3"
691..2084
/gene="SREBP3"
/codon_start=1
/product="sterol regulatory element binding protein 3"
/protein_id="AAF67491.1"
/db_xref="GI:7688701"
/translation="MEQKNVPPPEVCITHOGEKISANENSIAVRSPTAEDDRSDS
QVQVQVQVHPKPLSPDSRLSSPPKAKKFPAPARETCVCCQTVTPMERLL
AMQOVPHISCFRCYCNKSLISLTYSALHGRYICKPHNQLFKSGNYDEGFGHRPHK
DLWASKNEVEILERPAULANARETPHSQGVEDAPIAKVGLAASMEAKASQOEKEN
RAETTKKLIAMPPPTTELGSAGALEEGIKMSKPKWPPDEISKPEVPEDVLDLKL
RSSSLKERSRPFVVAASFQTSVKSPKTVSPPIKGMSECSQSESVGGRVAERKQV
ENAKASKNGNVKTTWQNKSKGTGKRSKEGHSLEMINENLVENGADSDDEDSFL
KQSPQEPKSLNWSFVDNTFAEFTTQNKSQDVLEWEGEVVVKELSVBEQIKRNY
DEDEDEE"

BASE COUNT 1052 a 620 c 703 g 973 t
ORIGIN

Query Match 62.7%; Score 2324; DB 9; Length 3348;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2694; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1001 AATGAGCTGAAGCCAGTGGTGGCAAAATCAAAATTCATAAAATGGAGCAAAAGGAGAT 1060
DB 649 AATGAGCTGAAGCCAGTGGTGGCAAAATCAAAATTCATAAAATGGAGCAAAAGGAGAT 708
QY 1061 GTGCCCCAGGTCTCTAGGTCTGCATCAACCATCAGGAAGGGAAAGATTCTTGCAAAT 1120
DB 709 GTGCCCCAGGTCTCTAGGTCTGCATCAACCATCAGGAAGGGAAAGATTCTTGCAAAT 768
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGATGACTCCCGTGAATCCAGGTT 1180
DB 769 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGATGACTCCCGTGAATCCAGGTT 828
QY 1181 AAGAGTGAAGTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
DB 829 AAGAGTGAAGTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 888
QY 1241 AGTCTTTCTGAAGTCTCTCCCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAG 1300
DB 889 AGTCTTTCTGAAGTCTCTCCCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAG 948
QY 1301 ACCTGGTGAAGTGCAGAGACAGTCTATCCAAATGAGAGGTTCTTTGGCCAAACAGAGAG 1360
DB 949 ACCTGGTGAAGTGCAGAGACAGTCTATCCAAATGAGAGGTTCTTTGGCCAAACAGAGAG 1008
QY 1361 GTGTTTCACATCAGTCTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACA 1420
DB 1009 GTGTTTCACATCAGTCTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACA 1068
QY 1421 TATGCACTTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
DB 1069 TATGCACTTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1128
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1540
DB 1129 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1188
QY 1541 AATGAAACGAAGATTTTGGAGACAGACCCAGCTTGCAAAATGCAAGGAGACCCCT 1600
DB 1189 AATGAAACGAAGATTTTGGAGACAGACCCAGCTTGCAAAATGCAAGGAGACCCCT 1248
QY 1601 CACAGCCAGGGGTAGAAGATGCCCTATTGTAAGGTGGTGTCTGCTGCTCAAGTATG 1660
DB 1249 CACAGCCAGGGGTAGAAGATGCCCTATTGTAAGGTGGTGTCTGCTGCTCAAGTATG 1308
QY 1661 GAAGCCAGGCTCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTG 1720
DB 1309 GAAGCCAGGCTCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTG 1368
QY 1721 AGGATCGCTGGCCACCCCTGAACTGGAAGTTTCAAGAGTGCCTTTGAGGAAGGG 1780
DB 1369 AGGATCGCTGGCCACCCCTGAACTGGAAGTTTCAAGAGTGCCTTTGAGGAAGGG 1428
QY 1781 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAGAGCAATCAGCAAGCCCAAGTCTCT 1840
DB 1429 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAGAGCAATCAGCAAGCCCAAGTCTCT 1488
QY 1841 GAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTTCTCACTGAAGCAAGAGAGC 1900
DB 1489 GAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTTCTCACTGAAGCAAGAGAGC 1548
QY 1901 CGCCCATTCAGTGAAGCTTCAATTTCAAAGCACTCTGTCAAGAGCCCAAAACTGTG 1960
DB 1549 CGCCCATTCAGTGAAGCTTCAATTTCAAAGCACTCTGTCAAGAGCCCAAAACTGTG 1608
QY 1961 TCCCCACCTTATCAGGAAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGGTGGA 2020
DB 1609 TCCCCACCTTATCAGGAAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGGTGGA 1668
QY 2021 AGAGTTGCAAGAGGAAACAAAGTGAAGATTCGAAGGCTTCTTGAAGCAATGGGATGTG 2080
DB 1669 AGAGTTGCAAGAGGAAACAAAGTGAAGATTCGAAGGCTTCTTGAAGCAATGGGATGTG 1728
QY 2081 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAGAAGTAAAGAA 2140

DB 1729 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGAGAGAGACAGGGAAGAGAGTAAAGAA 1788
QY 2141 GGTCTAGTCTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCCATGAA 2200
DB 1789 GGTCTAGTCTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCCATGAA 1848
QY 2201 GATGATAACAGCTTCTCTCAAAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTGG 2260
DB 1849 GATGATAACAGCTTCTCTCAAAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTGG 1908
QY 2261 AGTCTTGTAGACAACACCTTTGCTGAAGAAATTCAGTCTAGAAATTCAGAAATTCAGAGAT 2320
DB 1909 AGTCTTGTAGACAACACCTTTGCTGAAGAAATTCAGTCTAGAAATTCAGAAATTCAGAGAT 1968
QY 2321 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2380
DB 1969 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2028
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTTGCAATGATGCTGGCCCTT 2440
DB 2029 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTTGCAATGATGCTGGCCCTT 2088
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
DB 2089 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2148
QY 2501 GTATCCCAAGCATGAATGTAATTTACTTTGGAAGTAACTTTTGGAAAAAGAAATTCCTTCTTAA 2560
DB 2149 GTATCCCAAGCATGAATGTAATTTACTTTGGAAGTAACTTTTGGAAAAAGAAATTCCTTCTTAA 2208
QY 2561 AATCAAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAA 2620
DB 2209 AATCAAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAAACAACAAAA 2268
QY 2621 TTAATTTCTTCAATTTAGCAGTGATGATATGCATAAGTCTGTAAAGCTTGTAACTGGG 2680
DB 2269 TTAATTTCTTCAATTTAGCAGTGATGATATGCATAAGTCTGTAAAGCTTGTAACTGGG 2328
QY 2681 AAATATTCACCTGATTAATAGCCAGATTTCTGTAATCCCAAAAGGCAATATTAAGGT 2740
DB 2329 AAATATTCACCTGATTAATAGCCAGATTTCTACTGTATTCCTCAAAAGGCAATATTAAGGT 2388
QY 2741 AGATAGATGATTAGTATATTTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2800
DB 2389 AGATAGATGATTAGTATATTTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2448
QY 2801 GAATTTAGGGCTTTAAACATTTACGACTGAATGCACTTTTAGTATTAAGGGCACAGTTTGT 2860
DB 2449 GAATTTAGGGCTTTAAACATTTACGACTGAATGCACTTTTAGTATTAAGGGCACAGTTTGT 2508
QY 2861 TATTTTAAATGAATACCAATTTAAATTTTAGTATTTTACCTGTTAAAGAGATTTATTTAGT 2920
DB 2509 TATTTTAAATGAATACCAATTTAAATTTTAGTATTTTACCTGTTAAAGAGATTTATTTAGT 2568
QY 2921 CTTTAAATTTTAGGTTAAATTTCTTGTGTGATATATATAGAGAAATTTTACTATTTAT 2980
DB 2569 CTTTAAATTTTAGGTTAAATTTCTTGTGTGATATATATAGAGAAATTTTACTATTTAT 2628
QY 2981 GTCTGTCTCTCAACCTACATCTCGACTCGAGCTGAGGTATATATACACAGAGCAC 3040
DB 2629 GTCTGTCTCTCAACCTACATCTCGACTCGAGGTATATATACACAGAGCAC 2688
QY 3041 TTTTGGAGCAATTTGAAAAACCAACCTTACACTCTCTCGTCTTAGAGAGATCTGCTGCT 3100
DB 2689 TTTTGGAGCAATTTGAAAAACCAACCTTACACTCTCTCGTCTTAGAGAGATCTGCTGCT 2748
QY 3101 CCAAAATAGGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTTGCTTTCTT 3160
DB 2749 CCAAAATAGGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTTGCTTTCTT 2808
QY 3161 TCTGGT-GATATCTGCTCTCTCAATTAATTAAGTCTCAATATTTTAGTAAATACCTT 3219

Db	2809	TCTGGTGATATCTGTCTCTCATTAATCTGAAAGCTGCAATATTTTAGTAATACCTT	2868
Qy	3220	CGGATCACTGTCCTCCCATCTTCGGTGTAGAGCAAGTGAAGTTTAAAGGAGGA	3279
Db	2869	CGGATCACTGTCCTCCCATCTTCGGTGTAGAGCAAGTGAAGTTTAAAGGAGGA	2928
Qy	3280	AGAAAGAACTGCTTTACACCACTTGAAGTCTGAGCTGAGCTCTAAACCTGTAATTCCTTATGA	3339
Db	2929	AGAAAGAACTGCTTTACACCACTTGAAGTCTGAGCTGAGCTCTAAACCTGTAATTCCTTATGA	2988
Qy	3340	TGTCCTCTTTTGGAGACATAATTTTAAATACCTTACTAGCTCTGGAATATATGATTTT	3399
Db	2989	TGTCCTCTTTTGGAGACATAATTTTAAATACCTTACTAGCTCTGGAATATATGATTTT	3048
Qy	3400	TATCACAGTATTCAGGTTGAATTTAAACCACTATAGGCTTTTCTTGGAGATTTT	3459
Db	3049	TATCACAGTATTCAGGTTGAATTTAAACCACTATAGGCTTTTCTTGGAGATTTT	3108
Qy	3460	TCTAGTCTTAAAGTTTGGGACATATATAAATCTGAGTACATTTGTTGACACAGTTGATA	3519
Db	3109	TCTAGTCTTAAAGTTTGGGACATATATAAATCTGAGTACATTTGTTGACACAGTTGATA	3168
Qy	3520	TTCCAAATTTGATGATGGAGGAGAGGTGTCTTAAAGCTGTAGGCTTTCTTTGTACTG	3579
Db	3169	TTCCAAATTTGATGATGGAGGAGAGGTGTCTTAAAGCTGTAGGCTTTCTTTGTACTG	3228
Qy	3580	CATTATAGAGATTTAGCTTTAATTTTATAGAGTGTAAACATTTCTCTTTCTTAGT	3639
Db	3229	CATTATAGAGATTTAGCTTTAATTTTATAGAGTGTAAACATTTCTCTTTCTTAGT	3288
Qy	3640	CTTACCTAGTCTGAACATTTTATTTCAATAAGATTTTAAATTTGAAAAA	3699
Db	3289	CTTACCTAGTCTGAACATTTTATTTCAATAAGATTTTAAATTTGAAAAA	3348

RESULT 10
AC008147/C
LOCUS
DEFINITION Homo sapiens 12 BAC RP3-405J10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
AC008147
AC008147.31 GI:18464004
KEYWORDS HYG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 115345)

Munry,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amarutunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M., Davy-Carroll,L., David,R., Davila,M.L., Davis,C., Dey-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,F., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,K., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Liu,C., Liu,J., Liu,W., Louissegh,H., Lozard,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,

Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S., Umanli,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Zscherlupati,R., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 115345)
Worley,K.C.
Submitted (27-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115345)
Worley,K.C.
Direct Submission
Submitted (01-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 115345)
Worley,K.C.
Direct Submission
Submitted (28-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 115345)
Worley,K.C.
Direct Submission
Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Feb 1, 2002 this sequence version replaced gi:15626025.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu
CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.
ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.
SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

```
FEATURES
  source      Location/Qualifiers
              1..115345
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="12"
                /clone="RP3-405J10"
  misc_feature complement(2..2047)
                /notes="overlaps bases 1..2020 of clone AC139016"
                /function="clone overlap"
  repeat_region complement(2..79)
  repeat_region /rpt_family="AluY"
  repeat_region 110..156
  repeat_region /rpt_family="(A)n"
  repeat_region 157..470
  repeat_region /rpt_family="AluSg"
  repeat_region 471..484
  repeat_region /rpt_family="(A)n"
  STS          998..1132
                /standard_name="19465"
  STS          2114..2395
                /standard_name="133863"
  repeat_region 3365..3655
  repeat_region /rpt_family="AluJb"
  repeat_region 3668..4037
  repeat_region /rpt_family="LTR24"
  repeat_region 4041..4067
  repeat_region /rpt_family="(TA)n"
  repeat_region complement(4068..4242)
  repeat_region /rpt_family="AluSg/X"
  repeat_region complement(4243..4539)
  repeat_region /rpt_family="AluSc"
  repeat_region complement(4541..4836)
  repeat_region /rpt_family="AluY"
  repeat_region complement(4837..4890)
  repeat_region /rpt_family="AluSg/X"
  repeat_region complement(4891..5005)
  repeat_region /rpt_family="AluJo"
  repeat_region 5006..5118
  repeat_region /rpt_family="AluSx"
  repeat_region 5119..5425
  repeat_region /rpt_family="AluSx"
  repeat_region 5426..5585
  repeat_region /rpt_family="AluSx"
  repeat_region complement(5586..5690)
  repeat_region /rpt_family="AluJo"
  repeat_region 5931..6233
  repeat_region /rpt_family="AluY"
  repeat_region 8022..8320
  repeat_region /rpt_family="AluSg"
  repeat_region 8632..8934
  repeat_region /rpt_family="AluSg"
  repeat_region 9461..9759
  repeat_region /rpt_family="AluSx"
  repeat_region 9920..9947
  repeat_region /rpt_family="(TTTTG)n"
  repeat_region complement(9948..10229)
  repeat_region /rpt_family="AluSg"
  repeat_region 11773..11943
  repeat_region /rpt_family="LTR47A"
  repeat_region complement(14835..15109)
  repeat_region /rpt_family="AluSg"
  repeat_region complement(15112..15412)
  repeat_region /rpt_family="AluSg"
```

```
repeat_region complement(15467..15779)
repeat_region /rpt_family="AluSp"
repeat_region 15783..15818
repeat_region /rpt_family="(T)n"
repeat_region 15932..16230
repeat_region /rpt_family="AluJo"
repeat_region 16339..16637
repeat_region /rpt_family="AluJb"
repeat_region 16640..16933
repeat_region /rpt_family="AluY"
repeat_region 16934..16956
repeat_region /rpt_family="(CAAA)n"
repeat_region 16962..17249
repeat_region /rpt_family="AluSg"
repeat_region 17251..17288
repeat_region /rpt_family="AT-rich"
repeat_region 18079..18112
repeat_region /rpt_family="(TG)n"

Query Match      60.3%; Score 2234; DB 9; Length 115345;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2284; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1408 CAGTCTAGGAACATATGCATCTTTACATGGAAGATCTATTGTAAAGCCTCACTTCAATCA 1467
DB      |||||
DB 3172 CAGTCTAGGAACATATGCATCTTTACATGGAAGATCTATTGTAAAGCCTCACTTCAATCA 3113
QY      |||||
QY 1468 ACTCTTTAAATCTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCT 1527
DB      |||||
DB 3112 ACTCTTTAAATCTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCT 3053
QY      |||||
QY 1528 ATGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCNAATGC 1587
DB      |||||
DB 3052 ATGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCNAATGC 2993
QY      |||||
QY 1588 AAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGTAAGTGGGTGTCT 1647
DB      |||||
DB 2992 AAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGTAAGTGGGTGTCT 2933
QY      |||||
QY 1648 GCCTCAAGTATGAAGCCAGGCTCTCTCAGCAGGAGAGAGAACAGCAGCAGCTGA 1707
DB      |||||
DB 2932 GCCTCAAGTATGAAGCCAGGCTCTCTCAGCAGGAGAGAGAACAGCAGCAGCTGA 2873
QY      |||||
QY 1708 AACCAAGAAGCTGAGGATCGCTCGGCCACCCCTCACTGAACTTGAAGTTTCAGGAAGTGC 1767
DB      |||||
DB 2872 AACCAAGAAGCTGAGGATCGCTCGGCCACCCCTCACTGAACTTGAAGTTTCAGGAAGTGC 2813
QY      |||||
QY 1768 CTTGGAGAGGGATCAAAAATGTCAAAAGCCCAATGGCCTCTGAAAGCGAAATCAGCAA 1827
DB      |||||
DB 2812 CTTGGAGAGGGATCAAAAATGTCAAAAGCCCAATGGCCTCTGAAAGCGAAATCAGCAA 2753
QY      |||||
QY 1828 GCCGAAAGTTCTGAGGATTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTCACT 1887
DB      |||||
DB 2752 GCCGAAAGTTCTGAGGATTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTCACT 2693
QY      |||||
QY 1888 GAAGAAAGAGCGCCCATTCATCTGAGCAGCTTCATTTCAAGACCTCTGTCAAGAG 1947
DB      |||||
DB 2692 GAAGAAAGAGCGCCCATTCATCTGAGCAGCTTCATTTCAAGACCTCTGTCAAGAG 2633
QY      |||||
QY 1948 CCCAAAATCTGTCTCCACCTATCAGAAAGGCTGGAGCATGTGAGAGAGTGAAGA 2007
DB      |||||
DB 2632 CCCAAAATCTGTCTCCACCTATCAGAAAGGCTGGAGCATGTGAGAGAGTGAAGA 2573
QY      |||||
QY 2008 GTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAA 2067
DB      |||||
DB 2572 GTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAA 2513
QY      |||||
QY 2068 GAATGGGAATGTGGGAAACACCTGGCAAAACAAAGTCTAAAGGAGAGACAGGAA 2127
DB      |||||
DB 2512 GAATGGGAATGTGGGAAACACCTGGCAAAACAAAGTCTAAAGGAGAGACAGGAA 2453
QY      |||||
QY 2128 GAGAAGTAAGGAAGTCTAGTTTGGAGATGGAGATGCAATCTTTAGAAAATGGTGC 2187
DB      |||||
```

2452	DB	GAGNAGTAAGGAAGGTCATAGTTTGGAGATGGAGAAATGAGAAATCTTGTAAGAAAATGGTGC	2399
2188	QY	AGACTCCGATGAAGATGATAACAGCTTCCTCAAAACAACAATCTCCACAAGAACCCCAAGTC	2247
2392	DB	AGACTCCGATGAAGATGATAACAGCTTCCTCAAAACAACAATCTCCACAAGAACCCCAAGTC	2333
2248	QY	TCGAAATGGTCGAGTTTTGTAGACAAACACTTTGCTGAAGAATTCCTACTCTCAGAAATCA	2307
2332	DB	TCGAAATGGTCGAGTTTTGTAGACAAACACTTTGCTGAAGAATTCCTACTCTCAGAAATCA	2273
2308	QY	GAATCCCGAGTGTGAACTCTGGGAGGAGAGTGGTCAAAAGAGCTCTCTGTGGAAGA	2367
2272	DB	GAATCCCGAGTGTGAACTCTGGGAGGAGAGTGGTCAAAAGAGCTCTCTGTGGAAGA	2213
2368	QY	ACAGATAAAGAGAAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTCGCAAT	2427
2212	DB	ACAGATAAAGAGAAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTCGCAAT	2153
2428	QY	GATGCTGGGCTTTAAATTCATGTTAGTTAGCGAGCCATCGCCCTTTGTCAAAATGTGA	2487
2152	DB	GATGCTGGGCTTTAAATTCATGTTAGTTAGCGAGCCATCGCCCTTTGTCAAAATGTGA	2093
2488	QY	TGCACATAACGAGGTATCCACGAGATGAATGTAAATTTACTTGAAGTAACTTTTGGAAAAG	2547
2092	DB	TGCACATAACGAGGTATCCACGAGATGAATGTAAATTTACTTGAAGTAACTTTTGGAAAAG	2033
2548	QY	AATTCCTCTTAAATCAAAAAACAAAAACAAAAACAAAAACAACTCTAAATACTA	2607
2032	DB	AATTCCTCTTAAATCAAAAAACAAAAACAAAAACAAAAACAACTCTAAATACTA	1973
2608	QY	GAGATAACTTTACTTAAATTCCTCAATTTTAGCAGTGATGATATGCGTAAAGTCTGTAAGG	2667
1972	DB	GAGATAACTTTACTTAAATTCCTCAATTTTAGCAGTGATGATATGCGTAAAGTCTGTAAGG	1913
2668	QY	CTTGTAACGGGAAATATTCACACTGATATAATAGCCAGATCTACTGTATTCCTCAAAAG	2727
1912	DB	CTTGTAACGGGAAATATTCACACTGATATAATAGCCAGATCTACTGTATTCCTCAAAAG	1853
2728	QY	GCAATATTAAGGTAGATAGATTAAGTAGTATATTTTACACACTATTTTGGAAATAGA	2787
1852	DB	GCAATATTAAGGTAGATAGATTAAGTAGTATATTTTACACACTATTTTGGAAATAGA	1793
2788	QY	GAACATACAGAAGGAATTTAGGGGCTTAAACATTAAGCTGAATGCACCTTACTATAAAG	2847
1792	DB	GAACATACAGAAGGAATTTAGGGGCTTAAACATTAAGCTGAATGCACCTTACTATAAAG	1733
2848	QY	GGCACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTACTATTTACTCTGTAA	2907
1732	DB	GGCACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTACTATTTACTCTGTAA	1673
2908	QY	GAGATTATTTAGTCTTTAAATTTTTAGGTAAATTTTCTGTGTATATATATGAGGA	2967
1672	DB	GAGATTATTTAGTCTTTAAATTTTTAGGTAAATTTTCTGTGTATATATATGAGGA	1613
2968	QY	TTTACTACTTTATGCTGCTCTTAAACTTAACCTGAATCTGAGCGTCTGAGGTATAA	3027
1612	DB	TTTACTACTTTATGCTGCTCTTAAACTTACATCTGAACTCGACGTCTGAGGTATAA	1553
3028	QY	TACAACAGAGCACTTTTGGGCAATTTGAAAACCAACCTACACTCTTCGGTGCTTAGAG	3087
1552	DB	TACAACAGAGCACTTTTGGGCAATTTGAAAACCAACCTACACTCTTCGGTGCTTAGAG	1493
3088	QY	AGATCTGCTCTCCCAAAATAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAAT	3147
1492	DB	AGATCTGCTCTCCCAAAATAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAAT	1433
3148	QY	GATTGCTTTCTTTCTGGTGATATCTGTGCTTCTCTATAATTTACTGAAAAGCTGCAATTTT	3207
1432	DB	GATTGCTTTCTTTCTGGTGATATCTGTGCTTCTCTATAATTTACTGAAAAGCTGCAATTTT	1373
3208	QY	TAGTAATACCTTCGGGATCACTGTGCCCATCTTCGTCTGTAGAGCAAGTGAAGTTTT	3267
1372	DB	TAGTAATACCTTCGGGATCACTGTGCCCATCTTCGTCTGTAGAGCAAGTGAAGTTTT	1313

Qy	3268	AAAGGAGGAAGAAGAAAC	TGTCTTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTA	3327
Db	1312	AAAGGAGGAAGAAGAAAC	TGTCTTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTA	1253
Qy	3328	TTTCCCTTATGATGTC	CCCTTTTGAGACACTAAATTTTTAAATACTTACTAGCTCTGAAA	3387
Db	1252	TTTCCCTTATGATGTC	CCCTTTTGAGACACTAAATTTTTAAATACTTACTAGCTCTGAAA	1193
Qy	3388	TATATTGATTTTTATC	CACAGTATCTCAGGGTGAAATTAACCCCACTATAGGCCCTTTTTC	3447
Db	1192	TATATTGATTTTTATC	CACAGTATCTCAGGGTGAAATTAACCCCACTATAGGCCCTTTTTC	1133
Qy	3448	TTGGGATGATTTTTCT	AGTCTTAAAGGTTTGGGACATTAATAA	CTTGACATCTTCTTGT 3507
Db	1132	TTGGGATGATTTTTCT	AGTCTTAAAGGTTTGGGACATTAATAA	CTTGACATCTTCTTGT 1073
Qy	3508	ACACAGTTGATATCC	AAATTTGATGGATGGGAGGAGGTGTCCTTAAGCTGTAGGCTT	3567
Db	1072	ACACAGTTGATATCC	AAATTTGATGGATGGGAGGAGGTGTCCTTAAGCTGTAGGCTT	1013
Qy	3568	TTCTTTGTACTGCAT	TTATAGAGATTTAGCTTTAATATTTTTAGAGATGTAAACCATTC	3627
Db	1012	TTCTTTGTACTGCAT	TTATAGAGATTTAGCTTTAATATTTTTAGAGATGTAAACCATTC	953
Qy	3628	TGCTTTCTTAGTCTT	ACCTAGCTGAAACATTTTTATTCAATAAAGATTTTAATTAATAAT	3687
Db	952	TGCTTTCTTAGTCTT	ACCTAGCTGAAACATTTTTATTCAATAAAGATTTTAATTAATAAT	893
Qy	3688	TTGAA	3692	
Db	892	TTGAA	888	

RESULT 11
BC010664
LOCUS ...
DEFINITION Homo sapiens, clone IMAGE:3854371, mRNA, partial cds.

RESULT 11

BC010664
LOCUS

DEFINITION
ACCESSION

RECEIVED
VERSION
KEYWORDS

KEYWORDS
SOURCE

ORGANISM

REFERENCE

REFERENCES
AUTHORS

TITLE
JOURNAL

11

REMARK
COMMENT

BC010664 2121 bp mRNA linear PRI 12-JUL-2001
Homo sapiens, clone IMAGE:3854371, mRNA, partial cds.
BC010664
BC010664.1 GI:14715008

Homo sapiens (human)

Homo sapiens
Eukaryota: M

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

I (bases 1 to 2121)
Strausberg, R.

Direct Submissions Submitted (10/1/2010 - 10/1/2011)

Gene Collection (N
Institute 31 Gen

USA
Annapolis, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

cdDNA Library Preparation: Life Technologies, Inc.

CONA Library Arrayed by: IRI-M.A.G.E. Consortium (LEND)
DNA Sequencing by: Baylor College of Medicine Human Genomes

Sequencing Center
Center code: BCM-HGSC

Web site: <http://www.hgs>
Contact: villalon@bcm.tmc

Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S.

A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can

through the I.M.A.G.E Series. TRAX Plate: 1

This clone was selected for full length seq

passed the following selection analysis.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: b Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES		Location/Qualifiers	
source	1..2121	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/c1one="IMAGE:3854371"	
		/tissue type="Colon, adenocarcinoma"	
		/clone_id="NIH_MGC_65"	
		/lab_host="DH10B"	
		/note="vector: pCMV-SPORT6"	
CDS	<1..831		
	/codon_start=1		
	/product="Unknown (protein for IMAGE:3854371)"		
	/protein_id="AAH10664.1"		
	/db_xref="GI:114715009"		
	/translation="ARETPSPGVDEPAPIAKVGLVLAASMEAKASSQOEREDKPAETKK LRIAMPPTLGLSGGSLBEGIKMSKPWPPEDEISKPEVEDVDLDLKLRRSSSLK ERSPFTVAASFOSTSVKSPKTVSPPIRKQWMSSEOSSESVGGRVAFERKOVENAKSK KNGVKTTHQNKESKETKRSKEGHSLEMENENLVENGADSDSDNDFLKQSQPQE PKSLNSSFVDNTFAFEFTTQNKSQDVELWEGEVVKELSVSEBQIKRNRYYDEDEDE "		
BASE COUNT	697 a 376 c 448 g 600 t		
ORIGIN			
Query Match 53.1%; Score 1967; DB 9; Length 2121;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 2117; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1586	GCAGGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCCCTATTCTAGAGTGGGTGTC	1645
Db	1	GCAGGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCCCTATTCTAAGGTGGGTGTC	60
QY	1646	CTGGCTCAAGTATGGAAGCAAGGCCCTCTCTCAGCAGGAGAGGAACAAGCCAGCT	1705
Db	61	CTGGCTCAAGTATGGAAGCAAGGCCCTCTCTCAGCAGGAGAGGAGACAAGCCAGCT	120
QY	1706	GAAACCAAGAGCTGAGGATCGCGTGGCCACCCCACTGAACTTGGAGATTGAGAAAGT	1765
Db	121	GAAACCAAGAGCTGAGGATCGCGTGGCCACCCCACTGAACTTGGAGATTGAGAAAGT	180
QY	1766	GCCTTGAGAGAGGATCAAAATGTCAAGCCCAAAATGGCCCTCTGAGAGCAAAATCAGC	1825
Db	181	GCCTTGAGAGAGGATCAAAATGTCAAGCCCAAAATGGCCCTCTGAGAGCAAAATCAGC	240
QY	1826	AAGCCCGAAGTTCTGAGGATGTGATCTAGATCTGAAGAGAGTAAAGCATCTCTTCA	1885
Db	241	AAGCCCGAAGTTCTGAGGATGTGATCTAGATCTGAAGAGAGTAAAGCATCTCTTCA	300
QY	1886	CTGAAGGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAG	1945
Db	301	CTGAAGGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAG	360
QY	1946	AGCCCAAAATCTGTGCCACCTATCAGGAAAGGCTGGAGCATGTGAGAGCAGAGTGAA	2005
Db	361	AGCCCAAAATCTGTGCCACCTATCAGGAAAGGCTGGAGCATGTGAGAGCAGAGTGAA	420
QY	2006	GAGTCTGTGGGTGAAGAGTTGAGAGAGGAAACAAGTGAAGAAATGCCAGGCTTCTAAG	2065
Db	421	GAGTCTGTGGGTGAAGAGTTGAGAGAGGAAACAAGTGAAGAAATGCCAGGCTTCTAAG	480
QY	2066	AAGAAATGGGAATGTGGGAAAAAACAACCTGGCAAAACAAGAAATCTAAAGAGAGACAGGG	2125
Db	481	AAGAAATGGGAATGTGGGAAAAAACAACCTGGCAAAACAAGAAATCTAAAGAGAGACAGGG	540
QY	2126	AAGAGAGTGAAGAGGTGATAGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAATGGT	2185
Db	541	AAGAGAGTGAAGAGGTGATAGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAATGGT	600
QY	2186	GCAGCTCCGATGAGATGATACAGCTTCTCAACACAAATCTCCACAAGAACCCCAAG	2245
Db	601	GCAGCTCCGATGAGATGATACAGCTTCTCAACACAAATCTCCACAAGAACCCCAAG	660
QY	2246	TCCTGAATGGTCGAGTTTGTAGACACACACCTTTGCTGAAGAAATTCACCTCTCAGAAAT	2305

Db	661	TTCTTGAAATGGTCAGATTTTGTAGACAAACACCTTTGCTGAAGAAATTCACCTACTCAGAA	720
QY	2306	CAGAAATCCAGGATGTGGAACCTCTGGAGGGAGAGTGGTCAAAAGAGCTCTCTGTGGAA	2365
Db	721	CAGAAATCCAGGATGTGGAACCTCTGGAGGGAGAGTGGTCAAAAGAGCTCTCTGTGGAA	780
QY	2366	GAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGA	2425
Db	781	GAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGA	840
QY	2426	ATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGT	2485
Db	841	ATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGT	900
QY	2486	GATGACACATAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAA	2545
Db	901	GATGACACATAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAA	960
QY	2546	AGAAATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAACACATTTCTAAATAC	2605
Db	961	AGAAATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAACACATTTCTAAATAC	1020
QY	2606	TAGAGATAACTTTTACTTAAATTTCTTATTTAGCAGTGATGATATGATGATGATGATGATG	2665
Db	1021	TAGAGATAACTTTTACTTAAATTTCTTATTTAGCAGTGATGATATGATGATGATGATGATG	1080
QY	2666	GGCTTGTAACTGGGAAATATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAA	2725
Db	1081	GGCTTGTAACTGGGAAATATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAA	1140
QY	2726	AGGCAATATTAGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2785
Db	1141	AGGCAATATTAGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1200
QY	2786	GAGAACATACAGAGAAATTTAGGGCTTAAACATTACGACTGAAATGACATTTAGTATAA	2845
Db	1201	GAGAACATACAGAGAAATTTAGGGCTTAAACATTACGACTGAAATGACATTTAGTATAA	1260
QY	2846	AGGGCACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTAAATTTTAAATTTTAA	2905
Db	1261	AGGGCACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTAAATTTTAAATTTTAA	1320
QY	2906	AGAGATTTTGTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	2965
Db	1321	AGAGATTTTGTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	1380
QY	2966	AATTTACTACTTTATGTCCTGCTCTAAACTACATCTGAACTCGAGTCTCTGAGGTAT	3025
Db	1381	AATTTACTACTTTATGTCCTGCTCTAAACTACATCTGAACTCGAGTCTCTGAGGTAT	1440
QY	3026	AATCAACAGAGCAGCTTTTGGGCAATTTGAAACCAACCACTTCTCGTCTCTAG	3085
Db	1441	AATCAACAGAGCAGCTTTTGGGCAATTTGAAACCAACCACTTCTCGTCTCTAG	1500
QY	3086	AGAGATCTGCTCTCCCAATAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAA	3145
Db	1501	AGAGATCTGCTCTCCCAATAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAA	1560
QY	3146	ATGATTGCTTTCTTTTCTGTTGATATCTGCTCTCTCATATTAATCTGAAAGCTGCAATAT	3205
Db	1561	ATGATTGCTTTCTTTTCTGTTGATATCTGCTCTCTCATATTAATCTGAAAGCTGCAATAT	1620
QY	3206	TTTAGTAATACCTTCGGGATCACTGCTCCCATCTTCCGTTGTAGGCAAAAGTGAAGAT	3265
Db	1621	TTTAGTAATACCTTCGGGATCACTGCTCCCATCTTCCGTTGTAGGCAAAAGTGAAGAT	1680
QY	3266	TTAAAGGAGGAGAAAGAAAGTGTCTTACCACTTTGAGCTCAGACCTCTAAACCCCTG	3325
Db	1681	TTAAAGGAGGAGAAAGAAAGTGTCTTACCACTTTGAGCTCAGACCTCTAAACCCCTG	1740
QY	3326	TATTTCCCTTATGATGTCCTCTTTTGGAGACACTAAATTTTAAATTTTAAATTTTAAATTTT	3385

Db 1741 TATTTCCTTATGATGTCCTCTTTTGAGACACTAAATTTTAAATCTACTAGCTCTGA 1800

Qy 3386 AATATATTGATTTTATCACAGTATTTCTCAGGGTGAATTAACCAACTATAGCCCTTTT 3445

Db 1801 AATATATTGATTTTATCACAGTATTTCTCAGGGTGAATTAACCAACTATAGCCCTTTT 1860

Qy 3446 TCTTGGGATGATTTTCTAGTCTTAAAGTTTGGGACATTAATACTTGAGTACATTTGTT 3505

Db 1861 TCTTGGGATGATTTTCTAGTCTTAAACGTTTGGGACATTAATACTTGAGTACATTTGTT 1920

Qy 3506 GTACACAGTTCATATTCCTCAAAATTTGATGGATGGGAGGAGGTTCTTAAGCTGTAGGC 3565

Db 1921 GTACACAGTTCATATTCCTCAAAATTTGATGGATGGGAGGAGGTTCTTAAGCTGTAGGC 1980

Qy 3566 TTTCTTTTGTACTGATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAAACAT 3625

Db 1981 TTTCTTTTGTACTGATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAAACAT 2040

Qy 3626 TCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATCAATAAGATTTTAATTTAA 3685

Db 2041 TCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATCAATAAGATTTTAAATTTAA 2100

Qy 3686 ATTTGAAAAAATAAAAAA 3705

Db 2101 ATTTGAAAAAATAAAAAA 2120

RESULT 12

AK000057 2667 bp mRNA linear PRI 22-FEB-2000

LOCUS Homo sapiens cDNA FLJ20050 f.1s, clone COL00688.

DEFINITION AK000057

ACCESSION AK000057.1 Gi:7019894

VERSION oligo capping; f.1s (full insert sequence).

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (sites) Kawabata, A., Hiki, J., Kobatake, N., Inagaki, H., Ikeda, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, F., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2667) Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.

AUTHORS Direct Submission

TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

JOURNAL NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source Location/Qualifiers

1. .2667

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="COL00688"

/tissue_type="colon"

/clone_lib="COL"

/note="cloning vector pME18SFL3"

10. .1380

/note="unnamed protein product"

/codon_start=1

/protein_id="BAA30914.1"

/db_xref="GI:7019895"

CDS

/translation="MEQENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDSRDS QYKSEVQVHPKPLSPUSRASSSESSPPKAMFQAPARETCVSCQKTVYPMERLLA NOQVPHICFRCYSYCNKXLSYASLHRIYKPHFNQKFKSGNYDEGFQHPHK LWAENBEILERPAQLANARETSPGVEDAPIAKGVLAASMEAKASQOEKEDK PAETIKRLRTAMPPTPELSSGSALEEGIKMSKPWPPEDISKPEVPEVDVLDLKKLR RSSLKERSEPITVAASQSTSVKSPKTVSPPIKGMMSGQSESVESVQVAVRQVE NAKASKXNGNVKTTWKNKSKGRTKRSKEGHSLEMENENLVENGASDEDDNSFLK QSPQEPKSLNMSFVDTFAEFTTQNKQSQDVVELWEGEVVVELSVEBEQIKRNYDD EDEDEE"

BASE COUNT 865 a 512 c 570 g 720 t

ORIGIN

Query Match 51.4%; Score 1903; DB 9; Length 2667;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1276 GAAGTTTCAGGCACCTGCAAGAGAGACCTGCTGGAATCTCAGAGACAGTCTATCCCAT 1335

Db 240 GAAGTTTCAGGCACCTGCAAGAGAGACCTGCTGGAATCTCAGAGACAGTCTATCCCAT 299

Qy 1336 GGAGCGTCTCTTGGCCCAACAGCAGGTGTTTCCACATCAGCTGCTTCCGTTCCTCTATTG 1395

Db 300 GGAGCGTCTCTTGGCCCAACAGCAGGTGTTTCCACATCAGCTGCTTCCGTTCCTCTATTG 359

Qy 1396 CAACAAACAACTCAGTCTAGGAACATATGCATCTTTTACATGGAAGAATCTATTGTAAGCC 1455

Db 360 CAACAAACAACTCAGTCTAGGAACATATGCATCTTTTACATGGAAGAATCTATTGTAAGCC 419

Qy 1456 TCATCTCAATCAACTCTTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGCACAGACC 1515

Db 420 TCATCTCAATCAACTCTTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGCACAGACC 479

Qy 1516 ACACAAGGATCTATGGGCAAGCAAAAATGAAAACCAAGAGATTTTGGAGAGACAGCCCA 1575

Db 480 ACACAAGGATCTATGGGCAAGCAAAAATGAAAACCAAGAGATTTTGGAGAGACAGCCCA 539

Qy 1576 GTTTCGAAATGCAAGGGAGACCCCTCAGAGCCAGGGGTAGGAAGATGCCCTTATTGCTAA 1635

Db 540 GTTTCGAAATGCAAGGGAGACCCCTCAGAGCCAGGGGTAGGAAGATGCCCTTATTGCTAA 599

Qy 1636 GTTGGGTGTCCTGGTCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGGAAGAAGA 1695

Db 600 GTTGGGTGTCCTGGTCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGGAAGAAGA 659

Qy 1696 CAAGCCAGTGAACCAAGAGCTGAGGATCGCTGGCCACCCCTGACCTGACCTTGGAG 1755

Db 660 CAAGCCAGTGAACCAAGAGCTGAGGATCGCTGGCCACCCCTGACCTGACCTTGGAG 719

Qy 1756 TTCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAAGCCCAATGGCCTCTCTGAAGA 1815

Db 720 TTCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAAGCCCAATGGCCTCTCTGAAGA 779

Qy 1816 CGAAATCAGCAAGCCCGAAGTTCTTGGAGATGTCGATCTAGATCTGAAGAAGCTAAGACG 1875

Db 780 CGAAATCAGCAAGCCCGAAGTTCTTGGAGATGTCGATCTAGATCTGAAGAAGCTAAGACG 839

Qy 1876 ATCTTCTTCACTGAAGGAAGAAGCGGCCCATTCATCTGAGCAGTCTTCAATTTCAAGCAC 1935

Db 840 ATCTTCTTCACTGAAGGAAGAAGCGGCCCATTCATCTGAGCAGTCTTCAATTTCAAGCAC 899

Qy 1936 CTCTGTCAAGGCCCAAAACCTGTCCTCCACCTATCAGGAAGGCTGGAGATCTCAGA 1995

Db 900 CTCTGTCAAGGCCCAAAACCTGTCCTCCACCTATCAGGAAGGCTGGAGATCTCAGA 959

Qy 1996 GCAGAGTGAAGAGTCTGTGGGTGGAGAGTTCAGAGAGGAACCAAGTGGAAAAATGCCAA 2055

Db 960 GCAGAGTGAAGAGTCTGTGGGTGGAGAGTTCAGAGAGGAACCAAGTGGAAAAATGCCAA 1019

Qy 2056 GCGTCTTCAAGAGAGTGAAGTGTGGGAAAAACCAACCTGGCAAAACCAAGATCTTAAAGG 2115

Db 1020 GCGTCTTCAAGAGAGTGAAGTGTGGGAAAAACCAACCTGGCAAAACCAAGATCTTAAAGG 1079

Qy 2116 ACAGACAGGAGAGAGTAAGAGAGTCTAGTTTGGAGATGGGAATGAGATCTTGT 2175

1080	AGAGACAGGGAAGAGAAGTAAAGGAAGTCAATGTTTGGAGATGGAGAATGAGAATCTTGT	1139
2176	AGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTCAAAACAACTCTCCACA	2235
1140	AGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTCAAAACAACTCTCCACA	1199
2236	AGAACCCAGTCTCTGAAATTTGGTTCGAGTTTGTAGACAAACCTTTGCTGAAGAAATTCAC	2295
1200	AGAACCCAGTCTCTGAAATTTGGTTCGAGTTTGTAGACAAACCTTTGCTGAAGAAATTCAC	1259
2296	TACTCAGAAATCAGAAATCCCGAGATGTGGAACTCTGGAGGGAGAAGTGTGCAAGAGCT	2355
1260	TACTCAGAAATCAGAAATCCCGAGATGTGGAACTCTGGAGGGAGAAGTGTGCAAGAGCT	1319
2356	CTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGGAAGTG	2415
1320	CTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGGAAGTG	1379
2416	ACAAATTCGAATGATGTGGGCCCTTAAATTCATGTTAGTCTTAGCGAGGCACCTGCCCTTT	2475
1380	ACAAATTCGAATGATGTGGGCCCTTAAATTCATGTTAGTCTTAGCGAGGCACCTGCCCTTT	1439
2476	GTCAAAATGTGATGCACATAAGCAGGTATCCGAGCATGAAATGTAATTTACTTTGGAAGTA	2535
1440	GTCAAAATGTGATGCACATAAGCAGGTATCCGAGCATGAAATGTAATTTACTTTGGAAGTA	1499
2536	ACTTTGGAAAGAAATTCCTTCTTTAAATCAAAACAAACAAACAAACAAACAAACAAAC	2595
1500	ACTTTGGAAAGAAATTCCTTCTTTAAATCAAAACAAACAAACAAACAAACAAACAAAC	1559
2596	TTCTAAATCTAGAGATAACTTTACTTAAATCTTCTTTAGCAGTGATGATATGCATA	2655
1560	TTCTAAATCTAGAGATAACTTTACTTAAATCTTCTTTAGCAGTGATGATATGCATA	1619
2656	AGTCTGTGAGGCTTGTAACTGGGAAATATTCCACTGATATAGCCAGATTTCTACTG	2715
1620	AGTCTGTGAGGCTTGTAACTGGGAAATATTCCACTGATATAGCCAGATTTCTACTG	1679
2716	TATTTCCAAAGCAATATTAAAGTAGATAGATTAGTAGTAGTATGTTTACACACTAT	2775
1680	TATTTCCAAAGCAATATTAAAGTAGATAGATTAGTAGTAGTATGTTTACACACTAT	1739
2776	TTTGGAAATAGAGAACATACAGAGGAATTTAGGGCTTAAACATTACGACTGAATGCAAC	2835
1740	TTTGGAAATAGAGAACATACAGAGGAATTTAGGGCTTAAACATTACGACTGAATGCAAC	1799
2836	TTTAGTATAAAGGGCACAGTTTCTATATTTTAAATGAATACCAATTTAATTTTTAGTA	2895
1800	TTTAGTATAAAGGGCACAGTTTCTATATTTTAAATGAATACCAATTTAATTTTTAGTA	1859
2896	TTTACCTGTTAAGAGATTATTAGTCTTTAAATTTTTTAGGTTAATTTTCTGCTGTGAT	2955
1860	TTTACCTGTTAAGAGATTATTAGTCTTTAAATTTTTTAGGTTAATTTTCTGCTGTGAT	1919
2956	ATATATGAGAAATTTACTACTTTATGCTGCTCTTAAACTACATCCCTGAACCTGCACGT	3015
1920	ATATATGAGAAATTTACTACTTTATGCTGCTCTTAAACTACATCCCTGAACCTGCACGT	1979
3016	CCTGAGGTATATACAAACAGAGCACTTTTGGAGCAATTTGAAACCAACCTACACTCTT	3075
1980	CCTGAGGTATATACAAACAGAGCACTTTTGGAGCAATTTGAAACCAACCTACACTCTT	2039
3076	CGGTGCTTAGAGAGATCTGCTGCTCCCAAAATAGCTTTTGATCTGCCAGTGAATTTAC	3135
2040	CGGTGCTTAGAGAGATCTGCTGCTCCCAAAATAGCTTTTGATCTGCCAGTGAATTTAC	2099
3136	TGTACTCCAAATGATTGCTTTCTTTTCTGGTGATATCTGTGTTCTCATAAATTTACTGAAA	3195
2100	TGTACTCCAAATGATTGCTTTCTTTTCTGGTGATATCTGTGTTCTCATAAATTTACTGAAA	2159
3196	GCTGCAATATTTTGTGTAATACCTTCGGGATCACT	3229

Db	2160	GTGCAATATTTAGTAATACCTTGGGATCACT	2193
RESULT 13	BC001247		
LOCUS	BC001247	2164 bp	MRNA
DEFINITION	Homo sapiens, epithelial protein lost in neoplasm beta, clone		
ACCESSION	MGC:4969 IMAGE:3452714, MRNA, complete cds.		
VERSION	BC001247		
KEYWORDS	BC001247.1	GI:12654808	
SOURCE	MGC		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2164)		
TITLE	Strausberg R.		
JOURNAL	Direct Submission		
REMARK	Submitted (11-DEC-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapps-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcdexail@stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 4 Row: 0 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA q1: 70203350.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

Location/Qualifiers
1. 2164
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="LocusID:51474"
   /db_xref="Taxon:9606"
   /clone_accession="X64969 IMAGE:3453714"
   /tissue_type="Placenta, chorionic
   /clone_lib="NIH_MGC_10"
   /lab_host="DRI10B"
   /note="Vector: pCMV-SPORT6"
187. 1983

```

CDs

BASE COUNT
ORIGIN

```
Query Match          51.3%; Score 1901; DB 9; Length 2164;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 214; Conservative 0; Mismatches 0; Indels 3
```

QY 431 CACAGAGCAGACCATCTCTCTGTGTAAGTGAACGACGACGCTGTTCTGAGGCCAAAGCT 490
Db 1 CACAGAGCAGACCATCTCTCTGTGTAAGTGAACGACGACGCTGTTCTGAGGCCAAAGCT 60
QY 491 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCACCTCTGAGGCCCTCGTT 550
Db 61 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCACCTCTGAGGCCCTCGTT 120
QY 551 CAGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGAAAGT 610
Db 121 CAGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGAAAGT 180
QY 611 AAAAAATGGAATTTGTCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAAATCAGT 670
Db 181 AAAAAATGGAATTTGTCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAAATCAGT 240
QY 671 GAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAG 730
Db 241 GAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAG 300
QY 731 ATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGT 790
Db 301 ATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGT 360
QY 791 GCAAGTGAAGAGATCTCTGAAAAAGAGCTATTCTCTAGATGACCTGGAAATAGGCCCA 850
Db 361 GCAAGTGAAGAGATCTCTGAAAAAGAGCTATTCTCTAGATGACCTGGAAATAGGCCCA 420
QY 851 GGTGAGTTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGCAAAATCTGGAA 910
Db 421 GGTGAGTTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGCAAAATCTGGAA 480
QY 911 CTTCCAGCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACACGAGGAGCTGTG 970
Db 481 CTTCCAGCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACACGAGGAGCTGTG 540
QY 971 TCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCAGCTGTTGCGCAATC 1030
Db 541 TCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCAGCTGTTGCGCAATC 600
QY 1031 AAAATTCTAAATTTGAGCAAAAGAGAAATGTCGCCAGGTCCTGAGGTCCTGATCACC 1090
Db 601 AAAATTCTAAATTTGAGCAAAAGAGAAATGTCGCCAGGTCCTGAGGTCCTGATCACC 660
QY 1091 CATCAGGAAGGGAAAGATTCTTGCBAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 1150
Db 661 CATCAGGAAGGGAAAGATTCTTGCBAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 720
QY 1151 GCCGAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCC 1210
Db 721 GCCGAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCC 780
QY 1211 AAGCCACTAAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCAAGCA 1270
Db 781 AAGCCACTAAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCAAGCA 840
QY 1271 ATGAAGAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGNAATGTCAAGACAGTCTAT 1330
Db 841 AT--GAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGNAATGTCAAGACAGTCTAT 897
QY 1331 CCAATGAGGCTCTCTTGGCCAAACAGCAGGTGTTTCAATCAGCTGCTTCCGTTGCTCC 1390
Db 898 CCAATGAGGCTCTCTTGGCCAAACAGCAGGTGTTTCAATCAGCTGCTTCCGTTGCTCC 957
QY 1391 TATTGCAACAAACACTCAGTCTAGGAACATATGCACTTTTACATGGAAGAACTATTGT 1450
Db 958 TATTGCAACAAACACTCAGTCTAGGAACATATGCACTTTTACATGGAAGAACTATTGT 1017
QY 1451 AAGCTCACTTCAATCACTCTTTAAATCTAAGGCACTATGATGAAGGCTTTGGGCAC 1510
Db 1018 AAGCTCACTTCAATCACTCTTTAAATCTAAGGCACTATGATGAAGGCTTTGGGCAC 1077

QY 1511 AGACACACAAAGATCTATGGCAAGCAAAATGAAAAAGAGATTTTGGAGAGCCA 1570
Db 1078 AGACACACAAAGATCTATGGCAAGCAAAATGAAAAAGAGATTTTGGAGAGCCA 1137
QY 1571 GCCAGCTTGCATAATGCAAGGAGACCCCTCACACCCAGGGGTAGAGATCCCTCTATT 1630
Db 1138 GCCAGCTTGCATAATGCAAGGAGACCCCTCACACCCAGGGGTAGAGATCCCTCTATT 1197
QY 1631 GCTAAGGTGGTGTCTCTGGCTCAAGTATGGAAGCAAGGCTCTCTCTCAGCAGGAGAG 1690
Db 1198 GCTAAGGTGGTGTCTCTGGCTCAAGTATGGAAGCAAGGCTCTCTCTCAGCAGGAGAG 1257
QY 1691 GAAGCAACCCAGCTGAAACCAAGAGCTGAGGATCGGCTGCCCCACCCCACTGAACTT 1750
Db 1258 GAAGCAACCCAGCTGAAACCAAGAGCTGAGGATCGGCTGCCCCACCCCACTGAACTT 1317
QY 1751 GGAAGTTCAGGAAGTGCCTTTCAGGAGGAGGATCAAAATGTCAAAGCCCAATTTGGCTCT 1810
Db 1318 GGAAGTTCAGGAAGTGCCTTTCAGGAGGAGGATCAAAATGTCAAAGCCCAATTTGGCTCT 1377
QY 1811 GAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTA 1870
Db 1378 GAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTA 1437
QY 1871 AGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCATTTCACTGTAGCAGCTTCAATTTCAA 1930
Db 1438 AGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCATTTCACTGTAGCAGCTTCAATTTCAA 1497
QY 1931 AGCAGCTCTGTCAAGAGCCCAAAATCTGTGTCGCCACCTATCAGGAAAGGCTGGAGCATG 1990
Db 1498 AGCAGCTCTGTCAAGAGCCCAAAATCTGTGTCGCCACCTATCAGGAAAGGCTGGAGCATG 1557
QY 1991 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCGAGGAGGAAACAAGTGGAAAT 2050
Db 1558 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCGAGGAGGAAACAAGTGGAAAT 1617
QY 2051 GCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAAATCT 2110
Db 1618 GCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAAATCT 1677
QY 2111 AAAGGAGACAGGAGGAGAGAGTAAAGAGGTCTAGTCTTGGAGATGGAGATGAGAAAT 2170
Db 1678 AAAGGAGACAGGAGGAGAGAGTAAAGAGGTCTAGTCTTGGAGATGGAGATGAGAAAT 1737
QY 2171 CTTGTAGAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTCTCAACAACATCT 2230
Db 1738 CTTGTAGAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTCTCAACAACATCT 1797
QY 2231 CCACAAGAACCCCAAGTCTCTGAAATGGTGGAGTTCGAGCAACACCTTTCCTGAAGAA 2290
Db 1798 CCACAAGAACCCCAAGTCTCTGAAATGGTGGAGTTCGAGCAACACCTTTCCTGAAGAA 1857
QY 2291 TTCCTACTCTCAGAAATCAGAAATCCAGGATGTGAACTCTGGAGGGAGAGTGGTCAAA 2350
Db 1858 TTCCTACTCTCAGAAATCAGAAATCCAGGATGTGAACTCTGGAGGGAGAGTGGTCAAA 1917
QY 2351 GAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 2410
Db 1918 GAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 1977
QY 2411 GAGTGACAAATTCGAATGATGCTGGCCTTAAATTCATGTTAGTCTTACGAGCCACTGC 2470
Db 1978 GAGTGACAAATTCGAATGATGCTGGCCTTAAATTCATGTTAGTCTTACGAGCCACTGC 2037
QY 2471 CTTTGTCAAAATGTGATGCATTAAGCAGGTATCCAGCAGTGAATGTAAATTTACTTGG 2530
Db 2038 CTTTGTCAAAATGTGATGCATTAAGCAGGTATCCAGCAGTGAATGTAAATTTACTTGG 2097
QY 2531 AAGTAACTTTGAAAGAAATTCCTTTTAAATTCAAAAACAAAA 2574
Db 2098 AAGTAACTTTGAAAGAAATTCCTTTTAAATTCAAAAACAAAA 2141


```
RESULT 14
BD160117
LOCUS BD160117 2207 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160117
VERSION BD160117.1 GI:27865875
KEYWORDS JP 2002191363-A/14960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2207)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 14960 09-JUL-2002;
HELIOS RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/14960
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..2207
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 692 a 438 c 512 g 565 t
ORIGIN
Query Match 41.0%; Score 1519; DB 6; Length 2207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAGAT 1060
DB 638 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAGAT 697
QY 1061 GTGCCCCAGGTCCTGAGTGTGATCACCATCAGGAAGGGGAAAGATTTCTGCAAT 1120
DB 698 GTGCCCCAGGTCCTGAGTGTGATCACCATCAGGAAGGGGAAAGATTTCTGCAAT 757
QY 1121 GAGAATAGCTGCGAGTTCGCTCCACCCCTGCGGAAGATGACTCCCGTCACTCCAGGTT 1180
DB 758 GAGAATAGCTGCGAGTTCGCTCCACCCCTGCGGAAGATGACTCCCGTCACTCCAGGTT 817
QY 1181 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
DB 818 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 877
QY 1241 AGTCTTTTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 1300
DB 878 AGTCTTTTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 937
QY 1301 ACCTGCGTGAAGTGCAGAGAGAGTCTATCCCAATGGAGCGTCTCTTGCCCAACAGAG 1360
DB 938 ACCTGCGTGAAGTGCAGAGAGAGTCTATCCCAATGGAGCGTCTCTTGCCCAACAGAG 997
QY 1361 GTGTTTCAATCAGCTGCTTCCGTTGCTCTATTGCAACAACAACTCAGTCTAGGAACA 1420
DB 998 GTGTTTCAATCAGCTGCTTCCGTTGCTCTATTGCAACAACAACTCAGTCTAGGAACA 1057
QY 1421 TATGCATCTTTACATGGAGAGATCTATTGTGAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
```

```
1058 TATGCATCTTTACATGGAGAGATCTATTGTGAAGCCTCACTTCAATCAACTCTTTAAATCT 1117
1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGGATCTATATGGCAAGCAAA 1540
1118 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGGATCTATATGGCAAGCAAA 1177
1541 AATGAAACGAAGAGATTTTGGAGAGACCAGCTTGAATGCAAGGGAGAGCCCT 1600
1178 AATGAAACGAAGAGATTTTGGAGAGACCAGCTTGAATGCAAGGGAGAGCCCT 1237
1601 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGGTGGTGTCTTGGCTGCAAGTATG 1660
1238 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGGTGGTGTCTTGGCTGCAAGTATG 1297
1661 GAAGCCAAAGGCTCTCTCAGCAGAGAGAAAGAACAGCCAGCTGAAACCAAGAGCTG 1720
1298 GAAGCCAAAGGCTCTCTCAGCAGAGAGAAAGAACAGCCAGCTGAAACCAAGAGCTG 1357
1721 AGGATCGCTGGCCACACCCCTCACTGAACTTGAAGTTCAGGAAGTTCCTTGGAGGAAGG 1780
1358 AGGATCGCTGGCCACACCCCTCACTGAACTTGAAGTTCAGGAAGTTCCTTGGAGGAAGG 1417
1781 ATCAAAATGTCAAAAGCCCAATGGCTCTCAAGAGCAAAATCAGCAAGCCGGAAGTTCCT 1840
1418 ATCAAAATGTCAAAAGCCCAATGGCTCTCAAGAGCAAAATCAGCAAGCCGGAAGTTCCT 1477
1841 GAGGATGCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGAGCAAGAGC 1900
1478 GAGGATGCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGAGCAAGAGC 1537
1901 CGCCCAATCACTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960
1538 CGCCCAATCACTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1597
1961 TCCCACTATCAGAAAGGCTGGAGATGTCAAGAGAGTGAAGAGTGTGTGGGTGGA 2020
1598 TCCCACTATCAGAAAGGCTGGAGATGTCAAGAGAGTGAAGAGTGTGTGGGTGGA 1657
2021 AGAGTTGAGAGAGAGAACTAGTGGAAATGCCAGGCTCTTAAAGAAATGGGAATGTG 2080
1658 AGAGTTGAGAGAGAGAACTAGTGGAAATGCCAGGCTCTTAAAGAAATGGGAATGTG 1717
2081 GGAAAAACAACTGGCAAAACAAAGAAATCTAAAGAGAGACAGGAGAGAGAAAGTAAAGAA 2140
1718 GGAAAAACAACTGGCAAAACAAAGAAATCTAAAGAGAGACAGGAGAGAGAAAGTAAAGAA 1777
2141 GGTCAATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 2200
1778 GGTCAATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 1837
2201 GATGATAACAGCTTCCCTCAAAACAAATCTCCCAAGAACCCCAAGTCTCTGAATGGTGC 2260
1838 GATGATAACAGCTTCCCTCAAAACAAATCTCCCAAGAACCCCAAGTCTCTGAATGGTGC 1897
2261 AGTTTGTAGACAAACCTTTGCTGAAGATTTCACTACTCAGAAATCAGAAATCCAGAT 2320
1898 AGTTTGTAGACAAACCTTTGCTGAAGATTTCACTACTCAGAAATCAGAAATCCAGAT 1957
2321 GTGAACTCTGGAGGAGAGAGTGGTCAAGAGAGCTCTGTGGAAGAACAGATAAAGAGA 2380
1958 GTGAACTCTGGAGGAGAGAGTGGTCAAGAGAGCTCTGTGGAAGAACAGATAAAGAGA 2017
2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGA CAATTTGCAATGATGCTGGGCTT 2440
2018 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGA CAATTTGCAATGATGCTGGGCTT 2077
2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAGCAG 2500
2078 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAGCAG 2137
2501 GTATCCCAAGCATGAAATGTAAATTTACTTGAAGTAACTTTGGAAAAAGAAATTCCTTCTTAA 2560
```

Db 2138 GTATCCAGCATGAATTAATTTACTTTGGAAGTAATCTTGGAAAAGAAATCTCTTCTTAA 2197

Qy 2561 AATCAAAAAC 2570

Db 2198 AATCAAAAAC 2207

RESULT 15

AK023649

LOCUS Homo sapiens cDNA FLJ13587 fis, clone PLACE1009246, weakly similar

DEFINITION co POLLEN SPECIFIC PROTEIN SP3.

ACCESSION AK023649

VERSION 1.0

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Nagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, F., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T. NEDO human cDNA sequencing project

TITLE Unpublished

JOURNAL 2 (bases 1 to 2207)

REFERENCE Isogai, T. and Otsuki, T.

AUTHORS Direct Submission

TITLE Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

JOURNAL NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES

source

1..2207

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="PLACE1009246"

/tissue_type="placenta"

/clone_lib="PLACE1"

/note="cloning vector: pME18SFL3"

680..2053

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14625.1"

/db_xref="GI:10435635"

/translation="MEQENYVPGPEVCIHQEGEKISANENSIAVSTPAEDSDRS QVKEVQVFPKPLSPDRASSLSESPPMKKFQAPARETCVCCQKTYPMERLL ANQVHFISCRQYCNKLSLGTYSLSHGRYICKPHFNQLPKSKGNYDEGPHRPHK DLWASKNEEIIERPAQLANARETPSPGVEDAPIAKVGVLAASMEAKASOOEKED KPAETKKLIARPPPTLEGGSSALEGKMKPKWPPEDEIISKPEVEDVDLDKLI RRSLSKRSRPFVVAASFQSTSVKSPKTPPIRKGMSEQSESVGGRVARKGV ENAKSKNGNVGHTQNKESKGETSKRSKEGHSLENNENLVENGADSDSDNFI KQSPQEPKSLNWSFVNTFAEEFTTQNKQSDVELWEGEVKELSVBEQIKRNYI DEDEDE"

BASE COUNT 692 a 438 c 512 g 565 t

ORIGIN

Query Match 41.0%; Score 1519; DB 9; Length 2207;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1001 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGGAAT 1060

Db 638 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGGAAT 697

Qy 1061 GTGCCCCCAGGTCCTGAGGTCTGCATCACCACCCATCAGGAAGGGGAAAAGATTTCTGCAAAAT 1120

Db 698 GTGCCCCCAGGTCCTGAGGTCTGCATCACCACCCATCAGGAAGGGGAAAAGATTTCTGCAAAAT 757

Qy 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTGACTCCCAAGTTT 1180

Db 758 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTGACTCCCAAGTTT 817

Qy 1181 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCC 1240

Db 818 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCC 877

Qy 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAGAGAG 1300

Db 878 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAGAGAG 937

Qy 1301 ACCTGCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAG 1360

Db 938 ACCTGCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAG 997

Qy 1361 GTGTTTCAATCAGTGTCTTCCGTTGCTCTTATTCGCAACAACAACTCAGTCTAGGAACA 1420

Db 998 GTGTTTCAATCAGTGTCTTCCGTTGCTCTTATTCGCAACAACAACTCAGTCTAGGAACA 1057

Qy 1421 TATGCAATCTTTACATGGAAGAATCTATTTGTAAGCCCTCACTCAATCAACTCTTTAAATCT 1480

Db 1058 TATGCAATCTTTACATGGAAGAATCTATTTGTAAGCCCTCACTCAATCAACTCTTTAAATCT 1117

Qy 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACCAAGATCTATGGGCAAGCAAA 1540

Db 1118 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACCAAGATCTATGGGCAAGCAAA 1177

Qy 1541 AATGAAACAAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGGAGAGACCCCT 1600

Db 1178 AATGAAACAAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGGAGAGACCCCT 1237

Qy 1601 CACAGCCCAAGGGGTAGAAGATGCCCTATTTGCTAAGTGGGTGCTCTGCTGCTCAAGTATG 1660

Db 1238 CACAGCCCAAGGGGTAGAAGATGCCCTATTTGCTAAGTGGGTGCTCTGCTGCTCAAGTATG 1297

Qy 1661 GAGCCCAAGGCTCTCTCTCAGCAGGAGAGAGCAAGCCAGCTGAAAACCAAGAGCTG 1720

Db 1298 GAGCCCAAGGCTCTCTCTCAGCAGGAGAGAGCAAGCCAGCTGAAAACCAAGAGCTG 1357

Qy 1721 AGGATCGCTGGCCACCCCTCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAAGG 1780

Db 1358 AGGATCGCTGGCCACCCCTCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAAGG 1417

Qy 1781 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTCTGAAAGCGAAATCAGCAAGCCCAAGTTCTCT 1840

Db 1418 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTCTGAAAGCGAAATCAGCAAGCCCAAGTTCTCT 1477

Qy 1841 GAGGATGTGATCTAGATCTGAAAGAGCTAAGACGATCTCTTCTACTGAAGGAAGAGAGC 1900

Db 1478 GAGGATGTGATCTAGATCTGAAAGAGCTAAGACGATCTCTTCTACTGAAGGAAGAGAGC 1537

Qy 1901 CGCCCAATTCAGTGTACGAGCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960

Db 1538 CGCCCAATTCAGTGTACGAGCTTCATTTCAAGCACCTCTGTGTCAAGAGCCCAAAACTGTG 1597

Qy 1961 TCCCAACCTTATCAGGAAAAGCTGGAGCATGTTCAGCAGCAGTGAAGAGCTCTGTGGGTGGA 2020

Db 1598 TCCCAACCTTATCAGGAAAAGCTGGAGCATGTTCAGCAGCAGTGAAGAGCTCTGTGGGTGGA 1657

Qy 2021 AGAGTTGAGAAAAGGAAAACAAGTGGAAAATGCCAAGGCTTCTTAAGAAAGATGGGAATGTG 2080

Db 1658 AGAGTTGAGAAAAGGAAAACAAGTGGAAAATGCCAAGGCTTCTTAAGAAAGATGGGAATGTG 1717

Qy 2081 GGNAAAACAACCTGGGCAAAACAAGATCTTAAGGAGAGACAGGGAAGAGAAATGAAGAA 2140

```
Db 1718 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAAAGTAAGGAA 1777
QY 2141 GGTCTAGATTGGAGATGGAGATGAGATCTTTAGAAAAATGGTCCAGACTCCGATGAA 2200
Db 1778 GGTCTAGATTGGAGATGGAGATGAGATCTTTAGAAAAATGGTCCAGACTCCGATGAA 1837
QY 2201 GATGATAACAGCTTCCCTCAAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTGGTCG 2260
Db 1838 GATGATAACAGCTTCCCTCAAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTGGTCG 1897
QY 2261 AGTTTGTAGACAACACCTTTGCTGAAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 2320
Db 1898 AGTTTGTAGACAACACCTTTGCTGAAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 1957
QY 2321 GTGGRACCTCGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2380
Db 1958 GTGGAACCTCTGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2017
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCCCTT 2440
Db 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCCCTT 2077
QY 2441 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATSCACATAAGCAG 2500
Db 2078 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATSCACATAAGCAG 2137
QY 2501 GTATCCACGATGAATGTAATTTACTTGAAGTAACITTTGAAAAGAAATTCCTTCTTAA 2560
Db 2138 GTATCCACGATGAATGTAATTTACTTGAAGTAACITTTGAAAAGAAATTCCTTCTTAA 2197
QY 2561 AATCAAAAAC 2570
Db 2198 AATCAAAAAC 2207
```

Search completed: January 6, 2004, 20:07:01
Job time : 13116 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:44:24 ; Search time 20 Seconds

(without alignments)
3649.601 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSPPNRQWTSLSRVTA.....LSVERQIKRNYDEDEEE 759

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: pirl:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	6.3	537	T13564	microtubule-associ
2	239	6.1	916	A27864	neurofilament trip
3	232	5.9	734	B42880	nucleolus-cytoplas
4	222	5.7	1313	F96573	hypothetical prote
5	218.5	5.6	200	G84822	probable LIM-domai
6	218	5.6	189	T03400	transcription fact
7	217	5.5	219	S28507	nucleolar phosphop
8	216	5.5	990	I51618	myosin heavy chain
9	215.5	5.5	2139	T18296	hypothetical prote
10	214	5.4	1233	S56271	protein UNC-89 - C
11	213	5.4	6642	T29757	hypothetical prote
12	212.5	5.4	3488	T34418	LIM domain protein
13	212	5.4	211	T47915	probable transcrip
14	212	5.4	226	T02467	hypothetical prote
15	212	5.4	1020	T29108	transport protein
16	211	5.4	1790	S67593	myosin heavy chain
17	211	5.4	1961	A61231	transcription fact
18	210	5.3	172	T50694	hypothetical prote
19	210	5.3	199	T47716	probable heat shoc
20	207	5.3	852	T06310	neurofilament trip
21	206.5	5.3	298	T27209	protein F32E10.3 [
22	205.5	5.2	1871	D96796	hypothetical prote
23	205	5.2	849	S00030	hypothetical prote
24	204.5	5.2	990	H88733	probable heat shoc
25	204.5	5.2	1091	T34107	neurofilament trip
26	204	5.2	1110	I51116	protein F32E10.3 [
27	203.5	5.2	913	T52485	hypothetical prote
28	201	5.1	1658	S55101	neurofilament prot
29	201	5.1	2116	A26655	myosin heavy chain

ALIGNMENTS

RESULT 1

T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-537 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match	6.3%	Score 246;	DB 2;	Length 537;
Best Local Similarity	22.3%;	Pred. No. 0.00039;		
Matches 192;	Conservative 124;	Mismatches 330;	Indels 214;	Gaps 38,
QY	77	ENPGLGAESHDTSLRNSSTEIRHRADHPAEVTSAAAGAKADQEEQIHPSRLRSPEA	136	
Db	2600	ERPSPAESAKDAA--ESVEKSKDASRPPSVVESTKADSTKGDISPS--PESVLEGPKDD	2655	
QY	137	LVQ----GRYPHIK---DGEDLKDHS-----TESKMKMNCIGESRHE-----VE-----	173	
Db	2656	VEKSKESRPPSVSGASITGDSKDVSRPASVSVSKDEHDKAESRRRSIAKVESVIDEAG	2715	
QY	174	KSEISENTDASGKTEKYNV---PLNRLKMMFEKPEPTQTKI-LRAQSRASGRKISNSY	229	
Db	2716	KSDSKSSQSDQKDEKSTLASKEASRRSVSVESKDDAKESRPESVVASGEVPVRESK	2775	
QY	230	S-LDDLEIG-PGQ-LSSTTFDSEKNE--SRNLELPRLSETSIKDRMAKYCAA-----	277	
Db	2776	SPLSKDTSRPGSVSVESVTADEKSEQQSRRSVAESVKADTKKD--GKSQEAERPSVD	2833	
QY	278	-----VSKOSSSTNTYNELKASGGIKIHKMKENVPVGPVVCITHQGEKISA	327	
Db	2834	ELLXDDDEKQESRQISITGSHKAMSTMGDESFPMDKAD-KSEKPSRPESVAESIKHENTKD	2892	
QY	328	NENSLAVRSTPAEDSDRSQVKSEV---QQPVHPKPLS-----PDSRASSL	370	
Db	2893	EESPLGSR-----DSVAESIKSDITKGEKSPKPSKEVSRPESVVGSIKDEKAESRESV	2947	
QY	371	SESSPPKAMKKFOA--PARETC-----VECOKTVYPMERLLANQ	408	

Db 2948 AESVKPSSKDATSAPPSKESHSRVSGLKDEGDKTTSRRVSADSIKDEKSLVQS 3007
Qy 409 VFHISCFRCYCNKLSGLTYASLHGRYICKPH-----FNOLFK 447
Db 3008 -----ASRPESEASLKDAAAPSQTSRPESVTESVKDGKSPVASKEASRPASVAEN 3059
Qy 448 SKGNVDGFGHRP-----HKOLWASKNENEELERPAOLANARETPHSPGVE 494
Db 3060 AKSADSEKQRPESLPQSKAGSIKDEKSLKPLAGKDEAKSKESRSRESVAEQFP-----LV 3115
Qy 495 DAPIAKGVGLAASMEAKASSQOEKEDPAETKRLRIAWPPPTLGGSGSALREGIKWSK- 553
Db 3116 SKEVSRPASVAESV--KDEAKSKESPLMSKE-----ASRPASVAGSVKDEAKSKESRSR 3170
Qy 554 ----PKWP-PEDETSKP-----EVPEDVDLDLKLRRSS-----SLKERSPPFTVAAS 596
Db 3171 ESVAEKSPLSKESRSPASVAESVKDADKSKESRSRESAEKSPASVAES 3230
Qy 597 F-----QSTSVKSP-----KTVSPPIKSGMSQES-----EESVGGVRAERKQV 636
Db 3231 IKDEAKSKESRSRESVAESPLPSKESRPTSVAESVKDEAKSKESRSRDSVAESKPL 3290
Qy 637 ENAKASKNGVNGVTKWQKESKGETKRSK--EGHSELENNENLVENGADSDSDNSPL 694
Db 3291 ASKEASRP-ASVAESVQDEAKSKESRSRESVAESKPLAYKEASRPASVAESIKDEAKS 3349
Qy 695 KQSQPQ-----PKSLNWSFVDNTFAEFTTQKQSDVELMEGEVWKE 739
Db 3350 KEESRSRESVAESPLASKESRPTSVAESVKDEAKSKESRSRDSVAESKPLASKESRSP 3409
Qy 740 LSVEEQIKRNRNYDEDEEE 759
Db 3410 ASVAESV-----QDEAEK 3422

RESULT 2
A:27864
Nucleolus tripartite M protein - human
N:Alternate names: NF-M (medium) protein
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
C:Accession: A27864; A30157
R:Myers, M.W.; Lazarini, R.A.; Lee, V.M.Y.; Schlaepfer, W.W.; Nelson, D.L.
EMBO J. 6, 1617-1626, 1987
A:Title: The human mid-size nucleolus subunit: a repeated protein sequence and the
A:Reference number: A27864; MUID:87275853; PMID:3608989
A:Accession: A27864
A:Molecule type: DNA
A:Residues: 1-916 <MYE>
A:Cross-references: GB:Y00067; NID:G35045; PIDN:CAA68276.1; PID:G35046
R:Lee, V.M.Y.; Otvos Jr., L.; Carden, M.J.; Hollosi, M.; Dietzschold, B.; Lazarini, R.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 1998-2002, 1988
A:Title: Identification of the major multiphosphorylation site in mammalian nucleolus
A:Reference number: A30157; MUID:88158120; PMID:2450354
A:Contents: annotation; phosphorylation sites
C:Genetics:
A:Introns: 360/3; 402/2
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; phosphoprotein
F:615,628,641,654,680/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 6.1%; Score 239; DB 2; Length 916;
Best Local Similarity 22.1%; Pred. No. 0.0001;
Matches 180; Conservative 117; Mismatches 318; Indels 198; Gaps 39;

Qy 15 SLRVAKELSLVKNKSGAIVFISYQKAAEETNMKESNTE---NLQHFRIKGLTL 70
Db 205 ALRKDIEEASLVK-----VELDKYQSLQDEAVF--LRSNHEEVADLLAQIQASHIT 255
Qy 71 VLKKKNWPNGLGA-----EHTSLNSNTE---IHRADHPAEVTSMAAGAK- 117
Db 256 VERKDYLKTDISTALKEIRSOLESHSQNMHQAEEWPKCRYAKLITAAEQNKAIRSAKE 315

Qy 118 --ADQEEQIHPRS-----PLRSPPEAL-----VOGRYPHIKDGEDLKDHSTESKWMENCL 165
Db 316 ETAYRRLQKQSIKESVVRGKTESLERQLSDIERHNH-----DLSSYQDTIQOLENEL 370
Qy 166 GSRREVEKSEISENTDASGKIEKYNVPLNKLQMFKEGTEPTQTKILRAQSRASGRKIS 225
Db 371 RQTKEMAR-----HUREYQDULN-VKALDIEIAAYRKLLEGE-----TRFS 413
Qy 226 ENSYSLDDLEIGQGLSSSTTFDEKXNERRNLEPRL-----SETSIKDRMAKY 274
Db 414 TFAGSI-----TGPLYTHRPITISSIKIQTKVAPKLVQHKFVEEIIIEETKVEDEKSEM 469
Qy 275 QAAVSKQSSSTNYTNELKASGEIKIHKWKEKNVP-----PGPEVCIT 318
Db 470 EZAL-----TAITEELAAVMKEEKAEEPEAESEEAEEVAEAKKSPKATAPDEV--K 521
Qy 319 HOEGEK-----ISANENSLAVRSTPAEDSDSQKSEVQVPHKPLSPDSRASSLSE 372
Db 522 EEEGKEEEEGEEEEEDECAKSDQAEEGSEKSEKEEGEEOE---GETEAEAGE 578
Qy 373 SSPPKAMKKFOAPARETCVEQCKTVYPMERLLANQVPHISCFRCSCYNNKLSLGTAYSL 432
Db 579 EAEAEKKEVEKSEVATK-----EELVADAKVE----- 608
Qy 433 HGRYICKPHFNOLFKSGNYDEGFGHRPHKOLWASKNENEELERPAOLANARETPHSPG 492
Db 609 -----KPEAKSPVPKSPVEEK--CKSP-----VPSVPEEKSGSPVPKSPVEEKSGSP- 655
Qy 493 VEDAFIAKV---VLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTLGGSGSALREGI 549
Db 656 VPSVPEEKSGSPVSKSPVEEKAKSPVPK--SPVEEAKSK-----AEVG-KGEQKEEE 706
Qy 550 KMSKPKWPPDEI--SKPEVPEDVDLDLKLRRSSSLKER--SRPFTVAASFSTSVKSPK 606
Db 707 KEVK--EAPKEEKVEKKEPKDVP---EKKAESPVEEAEVAVTTIKSVKHUEKETK 762
Qy 607 TVSPPIRKWSMSQSEESVSGR--VAERKOVENAKASKK-----NGNVGKTTWQNKESK 659
Db 763 BEGKPLQ-----EKEKACGEGSEEGSDKGAKGRKEDIANGEV-----EGK 809
Qy 660 GTGKRSKEGHSLENNENLVENGAD---SDSDNSFLKQSQSPKPLSNWSSFFVDNTFA 716
Db 810 BEVEQETKSGSGREEKGVVNTGLDLSPADEKKGDKSEKVVVTKTVEKITSEGGDGA 869
Qy 717 BEFTTQN---QKSOD-VELMEGEVVKLSVEE 744
Db 870 TKYITKSVTVTKQVEEHEETFEELVSTKVKVEK 902

RESULT 3
B42680
nucleolus-cytoplasm shuttle phosphoprotein - rat
N:Alternate names: Nopp140 protein B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: B42680; A42680; S27890; S30510; S30511
R:Meier, U.; Blobel, G.
Cell 70, 127-138, 1992
A:Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542; PMID:1623516
A:Accession: B42680
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <MEI>
A:Cross-references: GB:M94288
A:Experimental source: clone pTW6
A:Accession: A42680
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'Q', 181-734 <ME3>
A:Cross-references: GB:M94287; NID:G205749; PIDN:AAAA1718.1; PID:G205750
R:Meier, U.; Blobel, G.
submitted to the ENBL Data Library, May 1992

Db 580 MKLK-----ESLKDKEEDLKNVTAETISSLREWEGSVLEKIEELS-----KYKES 623

QY 608 VSPPIKGMWSMEQSESVGRVARKQVENAKSKNGNKGKTTWONKSKSGKETGRKSK 667

Db 624 LVDKETLQSIQTAEELKGRGAHMKQIBELSTANASLVDEATKLSIQVQSESDLKEKE 683

QY 668 EGH-----SLMEHENIENVGAD--SDEDNSFLKQSQPOEPKSLNWSFVDNTFAEEFT 720

Db 684 AGYLKKIEELSVANESLADNVTDLQSIQVQESKDKREVAYLKIEELSVA NESLVDKET 743

QY 721 TONKSDQVELWEGEVVKEL-SVEEQIKRN 749

Db 744 KLQHDQEAELRGREASHLKKIEELSKEN 773

RESULT 5

G84822

probable LIM-domain protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 15-Jun-2001

C:Accession: G84822

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84822

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-200 <STO>

A:Cross-references: GB:A8002093; NID:2088643; PIDN:AAB95275.1; GSPDB:GN00139

C:Genetics:

A:Map position: 2

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

Query Match 5.6%; Score 218.5; DB 2; Length 200;

Best Local Similarity 36.5%; Pred. No. 0.00017;

Matches 46; Conservative 23; Mismatches 40; Indels 17; Gaps 4;

QY 387 RETVECOQTVYPMERLLANQOVPHISCFRCSCYCNKLSLGTYSALHGRYCKPHFNQLF 446

Db 7 QOKRACEKTVYVELLSADGIVHKACFCCKSHCKSLQLSNYSMMEGVYCRPHFSQLF 66

QY 447 KSKGNYDEGFGHRPHKDLWASKNEEILERPQAOLANARETPHSPGVEDA-----P 497

Db 57 KESGSFSKNF-QSPAKPL--TDKTPELNRTPSRLAGM-----PSGTDKCAICTKTVYP 118

QY 498 IAKGVV 503

Db 119 IEKTVV 124

RESULT 6

T03400

probable transcription factor SF3 - common tobacco

N:Alternate names: LIM-domain protein

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 15-Jun-2001

C:Accession: T03400

R:Shen, W.H.; Gigot, C.

A:Title: Sequence and analysis of the EMBL Data Library, February 1997

A:Description: LIM domain-containing protein of Nicotiana tabacum.

A:Reference number: 214931

A:Accession: T03400

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-189 <SHE>

A:Cross-references: EMBL:Y11002

A:Experimental source: strain Bright Yellow 2

C:Genetics:

A:Gene: SF3

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

C:Keywords: DNA binding

F:10-61/Domain: LIM metal-binding repeat homology <LIM>

Query Match 5.6%; Score 218; DB 2; Length 189;

Best Local Similarity 37.2%; Pred. No. 0.00017;

Matches 45; Conservative 24; Mismatches 44; Indels 8; Gaps 3;

QY 387 RETVECOQTVYPMERLLANQOVPHISCFRCSCYCNKLSLGTYSALHGRYCKPHFNQLF 446

Db 7 QOKRACEKTVYVELLSADGIVHKACFCCKSHCKSLQLSNYSMMEGVYCKPHFPEQLF 66

QY 447 KSKGNYDEGFGHRPHKDLWASKNEEILERPQAOLANARETPHSPGVEDAIAKVV 502

Db 67 KESGNFNKNF-QSPAK---SAEKLPELTRSPSKAAGMFSGTQEKCATCGKTAYPLEKVT 122

QY 503 V 503

Db 123 V 123

RESULT 7

S28507

transcription factor SF3 - common sunflower

C:Species: Helianthus annuus (common sunflower)

C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 15-Jun-2001

C:Accession: S28507; S37656

R:Baltz, R.; Domon, C.; Pillay, D.T.N.; Steinmetz, A.

A:Title: Submitted to the EMBL Data Library, February 1992

A:Description: Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger

A:Reference number: S28507

A:Accession: S28507

A:Molecule type: DNA

A:Residues: 1-219 <BAL>

A:Cross-references: EMBL:X64392; NID:G18818; PID:G18819

A:Experimental source: strain HA 401B

R:Baltz, R.; Domon, C.; Pillay, D.T.N.; Steinmetz, A.

A:Title: Plant J. 2, 713-721, 1992

A:Reference number: S37656; MUID:93258417; PMID:1302629

A:Accession: S37656

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168, 'E', 170-219 <BA2>

A:Note: 169-Lys was also found

C:Genetics:

A:Gene: SF3

A:Introns: 46/3; 79/1; 92/3; 122/3

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

C:Keywords: DNA binding; zinc finger

F:11-62/Domain: LIM metal-binding repeat homology <LIM1>

F:110-161/Domain: LIM metal-binding repeat homology <LIM2>

Query Match 5.5%; Score 217; DB 2; Length 219;

Best Local Similarity 35.8%; Pred. No. 0.00023;

Matches 48; Conservative 23; Mismatches 45; Indels 19; Gaps 4;

QY 379 MKKFOAPARETCVECOQTVYPMERLLANQOVPHISCFRCSCYCNKLSLGTYSALHGRYVC 438

Db 1 MKSGTGTQK-CTVCEKTVLVLDKLVANQVRVHKACFCCHCNSTLKLNSFNPSFGVYVC 59

QY 439 KPHFNQLFKSKGNYDEGFGHRPHKDLWASKNEEILERPQAOLANARETPHSPGVEDA-- 496

Db 60 RHFFDQLFKRTGSLKSPDGT-----KFKPERTFSQETQSAN-RLSSPFEGTRDKCN 111

QY 497 -----PIAKGVV 503

Db 112 ACAKIVYPIERVKV 125

RESULT 8

IS1618

nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: F51618; S57757
R: Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: F51618; MUID: 56019267; PMID: 7593294
A:Accession: F51618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:g8895920; PIDN:CAA61368.1; PID:g895921
C:Genetics:
A:Gene: xNopp180
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.5%; Score 216; DB 2; Length 990;
Best Local Similarity 19.6%; Pred. No. 0.0016;
Matches 152; Conservative 115; Mismatches 272; Indels 238; Gaps 29;

QY 20 AKELSLV-----NNKSSAIVEIFSKYQAAEETNMKKSENLENLSQHFRKGLTLVLKK 74
DB 28 AKFSKATGVKNDSCPTSLDIFSDWVKSPPD---AKKEKPPAN-----GLPKK 73

QY 75 KWENPGLGASHSDLSLNSSTETHRADHPAEVTSAAAGAKADQEEQIHPRSRLRSPP 134
DB 74 K-----SAKSSSSSDSSSE-----EDEPPAKRAQAPAGGKK-----PVKAVQPK 114

QY 135 EALVQGRYPHIKQEDLKDHSSTESK-----MENCLGSRHEVEKSEISENTDAS 184
DB 115 KAKSSS-----EISSDESSEETKPPKPPKPAQTPKVAAVKTPTKKAKSSSSSSSE 169

QY 185 GKIEKYNVPLNRLKMPFKGEPTQTKILRAQSASGOKISSENSYSLDLLEIGPGOLSSS 244
DB 170 DEAKKKQPIKV-----PKQAVVKAAGLASNGK---TADSSSESDDSPPAKTTAA 219

QY 245 TPDSKNSRNLEPLSLSTSIKDRMAKYQAAVSKSSSTNTNKLKASGGEIKIHKME 304
DB 220 TKTP-----PTKPATAKPOAKK---TAGKKSSSREDSSD---SSDEEQTKAKSK 263

QY 305 QKEN---VPPGEVCITHQEGEKISANENSLAVRSTPAEDDRSDQVKS-EVQQPVH-- 357
DB 264 PKPDVVSVPPTSV-----SKKTLTQPGTKAKPESSDSSSDSSDEEQPAKKA 312

QY 358 -----PKPLGPDSSRASSLSSPPKAMKFKQAPARETCVQCQTVVPMERLLAN 406
DB 313 KIYPKAAASAPPLAKKAAETSTDSSESSEDEK----- 347

QY 407 QQVFHISCFRCSCYCNKLSLGTYSALHGRYICKPHFNQLFKSGNYDEGFHGRPHKDLWA 466
DB 348 -----KSSVKGKAA-----PKK----- 361

QY 467 SKNENEILERPAAQLANARETPHSGVGDAPIAKVGVLAASMEAKASSQOEKEDKP-AET 525
DB 362 -----APA-APDAKSTPVAALKKAPAKKA---SSSSSDSSSNEETTTKPAKT 407

QY 526 KKLRIAWPPTELGSSGSALEEGIKMSKPKWPEDEISK-----PEPVEDVDLDLKL 578
DB 408 TPAKSAATPTSKTPTNGKATPTSKTFAKGTPTSTAKXSSSSSDSSSDEETTTKPA 467

QY 579 RRSLSLKRSRPTVAASFQSTSVKSPKTVSPP-----IRKGSMSQSESVGGRVA 631
DB 468 AKITPAKSAATPTSKTPTNSKATPTSKTTPAKGTPTKSAKKDSSSDSSSDEKKT 527

QY 632 ERQVENAKASKNGNVGKTTWQKSGKETGK---RSKEGHSLEMLNENLVENGADSD 688
DB 528 PAKRAAKTTTPAKD---AAKTTTPAKAATTPAKPAKASTPGQVPTKSSSSSDSSDSS 584

QY 689 DMSFLKQSQPQB-----KSLNWSFVDNTFAFEFTTQNK 725
DB 585 ED-----EKSSAKPAVKTTPGKATSPVAVKSPVPAKASSSSSDSSSDEETTTTK 637

RESULT 9

T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
A:Accession: T18296
R:Guillen, N.
submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: EMBL:L03534; NID:gi850912; PID:gi850913; PIDN:AAB48065.1
C:Genetics:
A:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 5.5%; Score 215.5; DB 2; Length 2139;
Best Local Similarity 20.5%; Pred. No. 0.0045;
Matches 174; Conservative 146; Mismatches 334; Indels 193; Gaps 38;

QY 7 NRRQWTSLSRVTAKELSLVNNKSSAIVEIFSKYQAAEE-----TMKKRSNTENLS 61
DB 1071 NELOOTQLKGETEKSLAAQ-----VAATKASDERDTLSQLENEKLTTKNLT 1119

QY 62 QHFRKGTL-----TVLKKXWNPGLGASHSDLSLNSSTETHRADHPAEVT-----SHA 112
DB 1120 K--TKADLEKXISGLKQDYEDLEDDKNKIEGLRNAQRKIKELDD-----EITKGADVQY 1173

QY 113 ASGAKADQEEQIHPRSRLRSPPAL---VOGRYPHIKQGE-----DLKDHS 156
DB 1174 LOKQKEEVESQI---AKQBEKEAIGNDVKNKEKTIKEKELEIQLSLOEKLDETEVEKEDA 1230

QY 157 ESKQME-----NCLGESRHEVEKSEIS-----ENTDASGI 187
DB 1231 EKKKXIEIEKEMKALQEEKENVESSKNSTEDKKKLEDNLKDTQKLDQMTADNEKLAKA 1290

QY 188 EKYNVPLNRLKMPFKGEPTQTKILRAQSASGOKISSENSYSLDLLEIGPGQJSSSTFD 247
DB 1291 KDLAQLENEVDQNEKA-VADAELIN-KKQAQSDKELNSLKALEALTKAKSVESKNKD 1348

QY 248 SEKNESRNLEPLSLSTSIKDRMAKYQAAVSKSSSTNTNKLK----- 293
DB 1349 SENEKAALSEEDQANE-----KLKNIQADLRKATADLQEAENEKKAEEVQAQRKLVADNK 1403

QY 294 -----SGGEIKIHKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDS 343
DB 1404 KMTKLEIEIKARDENTYKVENYEVK-----LKRKEADUEEANEENLDIEKKORMNKEK 1456

QY 344 RDSQVKSVEQOPVHPKPLSPDSRASSLSSESPPKAMKFKQAPARETCVQCQTVVPMERL 403
DB 1457 QVKKLEGELEK-----TKDLNAAIAEKDSITAKK-QSDA---DLEELNKTVEEHEDEV 1506

QY 404 LA--NQQVFHISCFRCSCYCNKLSLGTYSALHGRYICKPHFNQLFKSGNYDEGFHGRPH 461
DB 1507 VAKLNTQI-----TKLTRDNQSA-----EEELNEL-RSKADKDK----- 1539

QY 462 KDLWASKNENEILERPAAQLANARETPHSGVGDAPIAKVGVLAASMEAKASSQOEKEDK 521
DB 1540 KKISELEEQVNELESRPVTGTGNADN---NEIKIRDAQIADLN-KALEMKGVQNNQLQATNK 1596

QY 522 PAE-----TKKLRIAWPPTELGSSGSALEEGIKMSKPKWPEDETSKBPVDPDVL 575
DB 1597 ELKAKNDLTSKIETENEMKKLENKAKRIEOD-KDPAKAVSQTTIKRKGLEEV----- 1651

QY 576 KKLRRS-SSLK-ERSRPTVAASFQSTSVKSPKTVSPPIRKGSMSQSESVSGRVAER 633
DB 1652 KKLTTETQALKFQINAPSSVAQEEKQRLS-----DIAELKEQLQERTTAANAER 1705

QY 634 KOVENAKSKNGVNGKTTWQ-----NKESKGETGKRSKEGHSLEMENLV-ENGAD 685
DB 1706 KKIQ-AELDEVKFNLEDVTNQREXVAKNGENDAEIDSLKEEKALEDEIEKITDNNKL 1764
QY 686 SDEDDNSELKQSQFQPSLWNS---SFVNTFAEBEFTTQKQSQDVLMW--EGSVVKEL 740
DB 1765 SEEDSLDRKYNALLDSKDSVMKEXFQDELKVTKOALETEKKNHAETMRLKGRLEKE- 1823
QY 741 SVEEQIK 747
DB 1824 AAEVOVR 1830
RESULT 10
S56271
hypochemical protein YFR016c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
C:Accession: S56271
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasano
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce
A:Reference number: S56186
A:Accession: S56271
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1233 <MUR>
A:Cross-references: EMBL:D50617; NID:9836685; PIDN:BA09255.1; PID:d1009896; PID:g836771
C:Genetics:
A:Cross-references: SGD:S0001912
A:Map position: 6R
Query Match 5.4%; Score 214; DB 2; Length 1233;
Best Local Similarity 20.5%; Pred. No. 0.0027;
Matches 183; Conservative 125; Mismatches 308; Indels 278; Gaps 46;
QY 43 KAAEETNMKRSN-TENLSQFRKG-----TLTVLKKWENPOLGAESHTDS 89
DB 334 KDVSSESLTKNGFNKFNKESKHLKAGEKQOTESDRDGISPSVLAKNQKETEIGKEDHVE 393
QY 90 LRNASTE-----IRHADHPPAEVTSHAASGAKAD-----OEEQIHPRSRL 130
DB 394 QKXEDERCKELSVNHENN-----NSHNFAAGSDSIIPPERETYDDTGMPTKRI 447
QY 131 RSPPEALVQR-----YPHIKOGEDLKHSTESK-KWNCLEGS-----RHEV---EKS 175
DB 448 SDNEKNLQHGNDISVEVEKEEBEENSTFSKVENVTGCEQFAVRNNEVSGTEEE 507
QY 176 EISENTASGKIEKYNVPLNLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDLE 235
DB 508 STSKGEEIMGDEKQS-----PAGE--KSIIEIEG-SANSAKISKONLVLEDEA 554
QY 236 IGPQLSSST-----FD-----SEKNESRRNLPL-----RLSE--T 265
DB 555 EAPTOENKPTVEGIDIPADPDVVEIVEAVENKIIPEDLEVAKEQGEQVQLDEPVK 614
QY 266 SIKRMAYQAVS-----KQSSNTYNEL-----KASGGEIKHKMQKE 307
DB 615 AMKDDKIARGAESISEDMKKQGTALSLNSKAKKEVDETARESAGVEV-----EKS 668
QY 308 NVPPQPEYVC--ITHQEGEKISANENSLAVRSTPAEDDSR--DSOVKSEV----- 352
DB 669 KTPESPKVVKCTSGRPDLQINERDPV---LKEDVRVDPEDVKPEIATINSEED 724
QY 353 -----QOPVHPKPLSPDRASSLSSESPPKAMKFKQAPARETCVQCQTVYPM 401
DB 725 PKSORVOISTEQAEITQKMDGVDGTTSTFKEEKPKRFEITQEGDKITGD----- 775
QY 402 RLLANQVPHISCFRCSCVNNKLS-LGTYASLHGRIRYCKPHFNGLFKSGKNYDEFGHRP 460
DB 776 ---TNHE--HGEATEAAGENSKSDVGT-----AEKVIEPSSSEVKD-----T 814
QY 461 HKDLWASKNENEIILERAQLANARETHPSGV-----EDAPI----- 498

DB 815 BEDAEVENSEKTEF:KVAELENL-DAPKEAVTAAELNKENEDVEDVTEDEAVENSEKT 873
QY 499 -----AKGVLAASMEAKASSQOEKPKAPETKRLRIAWPPPTTELGSSGSALEEGIKMS 552
DB 874 EPIKVAELGNLDAKAEAVTAAELNKENEDV-----AATSKEDIETKCS 919
QY 553 KPKWPP-----EDEISKPE-----VPEDVDLDLKKLRSSSLK-----ERSRPTVAAS 596
DB 920 EPAETPIEDGTCTEAEVSKKDAEAVTKEDENWENSKI--AEALKDVTGQOEIDINISDE 977
QY 597 POSTSVKSPKTVSPPIRKQWMSVSESESVGGRVRAERKQVENAKASKNGNVGKTTWQ-- 655
DB 978 FO-----RTVELP-----ELEKQIDKDKGDEKELEVEETEKETSLPDLVVEENI 1022
QY 656 KSKGETGKRSKEGHSLEM-ENENLVENCADSDDEDDNSELFLKQSQSQEPKSLN-----WSSF 710
DB 1023 TEENEIKOEVEEVSQLDFNETESISKEAPNDE--NGFDDQSTRENPKKASADDIFKDI 1080
QY 711 VONTFAEBEFTTQKQSQDVLMWEGVWKLVSVE-----EOIKNRYVDEDED 757
DB 1081 LDET--NEFLEQLKIVDDSEL--NALLQSLDAKDSITTQTEQSKNN--DKPD 1128
RESULT 11
T29757
protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C:Accession: T29757
R:Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C09D1.
A:Reference number: Z20679
A:Accession: T29757
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6642 <DUZ>
A:Cross-references: EMBL:AF003131; PIDN:AA54132.1; GSPDB:GN00019; CESP:unc-89
A:Experimental source: strain Bristol N2; clone C09D1
C:Genetics:
A:Gene: CESP:unc-89
A:Map position: 1
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;
3/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1
Query Match 5.4%; Score 213; DB 2; Length 6642;
Best Local Similarity 21.3%; Pred. No. 0.024;
Matches 178; Conservative 97; Mismatches 325; Indels 234; Gaps 36;
QY 21 KELSUNKNKSSAIVEIFSKYQAAAEETNMKRRNTENLSQHFRGTLTVLKKWENPG 80
DB 1085 KQSDRVEIREFDGSIKISIKNIKIEDAGEIRAVATNSEGSD--TKAKLTVQKKPF--- 1138
QY 81 LGAESHTDSLNSSTEIRHRAHPPAEVTSHAASAKADQOEQIHPRSLRSPPEALVQ 140
DB 1139 ---APEFLRPSVLTVEKSE---AVSAHAFGIPLPTYEWSVNGR-KVRDQGEARVT 1190
QY 141 RYPHIKDGEDL-----KDHSTESKKNENCLG-----ESRHE---VEKSE 176
DB 1191 RRESTDGASILITDATTYSEVNHLTISVVAENTLGAEETGAQLTIEPKKESVVEROD 1250
QY 177 ISENTDASGKIEKYNVPLNLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDLE 212
DB 1251 LS-----SSEVQK-----EIAQVKEASPEATTTITMETSLSITKTTTMSITEVTSTVG 1299
QY 213 -----RAQSRAS-----GRKISSENSYSLDLEIGPGQLSSSTTFSEKNESSR--- 255
DB 1300 GVTVEKSESESESATTVIGGSGGVTEGSIKIEV-VSKTDSQTDVREGTPKRVSPA 1358
QY 256 NLELPELSETS:KDRMAYQAAVSKQSSSTNTYNELKAS-----GGEIKHWEKENV 310
DB 1359 EELPK--EVIDSDRKKKSPSPDKKESPEKTEPKPASPTKKTGEEVK-----SPKESKP 1412

QY 311 PQEVCITHOGEKISANENSLAVRSTPAEDDSRDSQVSEVQVQVPHPKPLSPDSR-ASS 369
DB 1413 ASP-----TKKESPAEE-----VKSPTKKKESPSPTKKE-----KSPSPTKKTGDE 1457
QY 370 LSESSPKA-----MKFOAPARETCVE--CQKTVYPMERLLANQOVF 410
DB 1458 VKEKSPKSPKTKESPEKPEVDPVKSVPKKEKSPDATNIVEVSSETTIEKTTMTTEN-- 1515
QY 411 HSCFRCSYNNKLSLGTYSALHGRYCKPHFNQVFKSGNVDGFGHPRPHKDLWASQVE 470
DB 1516 -----THESBESRTSVKKE-----KTPKVDKPKSPKTKKDXSPEKSI 1553
QY 471 NBEIL-----ERPAQLANARETHSPG-----VEDAPIAKVG 502
DB 1554 TBEIKSPVKKEKSPKVEKPAKSPKTKESPKPASPTKKSENVKSPKTKKESKSP--EKS 1611
QY 503 VLAASWEAKASSQOEKEDAP-AETKKLRITAMP-----PTELGSSGSALDE-- 547
DB 1612 VVEELKSPKESPEKADDPKSPKTKK-----SPEKSATEDVKSPTKKESPEKVEEKTS 1668
QY 548 -GKMSKPKWPPDEIRSKPEVPEVDLDLKKLRRSSSLKERSRPTVAASFQSTSVKSPK 606
DB 1669 PTKKESPTKTDDEKSPKTKKESKSPQTVKPKPASPTKKKESPEKSVV-----BEVKSPK 1723
QY 607 TVSP-----PIRKGWSMEQSEESVGGRAERKQVENAKKNGNNGVCKTTWQNK 657
DB 1724 EKSPEKAEKSPKSPKTKKESPEKSAEYKSPKTKKESPEKSAEYKSPKTKKESPVKM 1783
QY 658 SKGETCKRSGEHSLE-----MENENLVENGADSDDEDNSFLKQOSPOEP 702
DB 1784 ADDEVKSPKTKKESPEKVEKPAKSPKTKKETPEKSA-ABELKSPKTKKESKSPSP 1836
RESULT 12
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP.F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1
Query Match 5.4%; Score 212.5; DB 2; Length 3488;
Best Local Similarity 19.5%; Pred. No. 0.012;
Matches 155; Conservative 129; Mismatches 328; Indels 183; Gaps 27;
QY 39 SKYQKAAETNMEKRSNTENISQHFRTKGTTLVLKKKNWPNGLGAESHTDSLURNSTEIR 98
DB 716 SKLEKAADTT---KQIETETVDD--KSKKKVKKKTE-----KSDSPTS 755
QY 99 HRADHP-REVTSHASGA-----KADQEQIHRP-----SRLRSPPEA 136
DB 756 QKSETPVVVEPTKPAESBAQKTAENVKAKKQKQEVDDNLKREAEVAAKKIADKLEKIAEA 815
QY 137 LVQGRYPHFKDGBDLKDHSTESKMKMNCGLGSRHVEKSEIGENTD---ASGKIEKYNV 192
DB 816 NIK-KTAEVEAAKKQE-KDEQLKLETEVTVKSKAAEKLKLEKQAKKAAADAVKQK 873
QY 193 PLNRLKXMFKEKGPPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGLSSSTFDSEKNE 252

DB 874 ELNEKNKLEAAKSAADKLEBESAAKSKVSESVKFEKKTKAGETKVQVESEPT- 932
QY 253 SRNLELPRLSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGEIKIHMEQKENVPPG 312
DB 933 SKTIDTKDVGATPEPADETTPKKIIKKKTEKSSSIQKSAITDSE-KVSQKQODE-PTK 990
QY 313 PEVCITH-----OGEKISANENSLAVRSTPAEDDSRDSQVSEVQVQVPHPKPLSPDSRA 367
DB 991 PAVSETQMTVEADKSKKQKQKDEKLDLDAELIAKTKQEADEKSKLDAQEKIKKVSDEDDAA 1050
QY 368 SSJSE-----SSPPKAMKFOAPARETC-VECOKTVYPMERLLANQOVFH 411
DB 1051 RKKEKLDKLESEIATKKASADKLEBQAQAKKAAVEAAK----- 1094
QY 412 ISCFRCYNNKLSLGTYSALHGRYCKPHFNQVFKSGNVDGFGHPRPHKDLWASQVEN 471
DB 1095 ---KQEKDEQLKLDTEAA-----SKKAAAEKLE- 1120
QY 472 EEILERPQALANA-----RETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDK- 521
DB 1121 ---LEKQAOIKKAAGADAVKKQKELDEKKNLEANKKSAAGLKIEBESAAKSKQTVBEOA 1177
QY 522 --PAETKKLRITAMPPTTELGSSGSALBEGIKNGKPKWPPDEISKPEVPEVDLDLKKLR 579
DB 1178 KLDATK-----AKTAEKQTKLEKDEKSTKES---ESKETVDEKPKKKVKKKTEK 1225
QY 580 RSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRK-----GHSMSQSE 623
DB 1226 SDSSISQKSETKTVBESAGPSESETQKADAAARKQKQKLEAEITAKKSADEKSK 1285
QY 624 ESVGGRVAERKQVENAKKNGNNGVCKTTWQNKESKQETGKRSKE--GHSLEMENENLVE 681
DB 1286 LEABSKLKAAEVEAAKKQE-----KDEQLKLDTEAAKSKAAAEKLKLEKQSHIK 1336
QY 682 NGADSDDEDNSFLKQOSPOPKSLNWSFVNDTFAEEFTTQNKOSQDVLEWEGEVWVWELS 741
DB 1337 KAAEVD---AVKQKLEBKQRLSEPAATKADAELKLEBQKKKAAETALIEQKE-- 1390
QY 742 VEEQIKERNRYDEDE 756
DB 1391 -QEKLAQEQSRLEDE 1404
RESULT 13
T47915
LIM domain protein - Arabidopsis thaliana
N:Alternate names: protein T20K12.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 15-Jun-2001
C:Accession: T47915
R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quer
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z24480
A:Accession: T47915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <DEH>
A:Cross-references: EMBL:AL137898
A:Experimental source: cultivar Columbia; BAC clone T20K12
C:Genetics:
A:Map position: 3
A:Introns: 46/3; 80/1; 89/3
A:Note: T20K12.130
C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology
Query Match 5.4%; Score 212; DB 2; Length 211;
Best Local Similarity 26.6%; Pred. No. 0.00039;
Matches 47; Conservative 29; Mismatches 47; Indels 54; Gaps 2;
QY 371 SESSPPKAMKFOAPARETCVECOKTVYPMERLLANQOVPHISCFRCYNNKLSLGTYA 430
DB 86 SDAAPNRLSSFFSGTQDKCAACKKTVVPLEKMTMEGESYHKTCFCAHSGCPLTHSSYA 145

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:46:24 ; Search time 42 Seconds
(without alignments)

4663.379 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSPFRQWTSLSLRVTA.....LSVEEQIKRNYDEDEEE 759

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 23:**

- 1: sp_archea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mmc:
- 8: sp_organellie:
- 9: sp_phage:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_rvirus:
- 16: sp_bacteriap:
- 17: sp_archeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2970	75.6	753	11	Q8k2h0 mus musculu
2	2955	75.2	753	11	Q8c7s2 mus musculu
3	2395	61.0	593	11	Q8cd09 mus musculu
4	2390	60.9	593	11	Q8bt15 mus musculu
5	2384	60.7	593	11	Q8c3r7 mus musculu
6	1128	28.7	629	13	Q9dey8 brachydanio
7	641	16.3	951	6	Q8hxj9 macaca fasc
8	529	13.5	519	4	Q8n7z0 homo sapien
9	363	9.2	128	11	Q8gob5 mus musculu
10	357	9.1	127	4	Q9bt23 homo sapien
11	357	9.1	127	4	Q96s91 homo sapien
12	258	6.5	2081	10	Q9lh98 arabidopsis
13	251	6.4	5412	5	Q9w596 drosophila
14	246	6.3	5327	5	Q76891 drosophila
15	241.5	6.1	7210	5	Q9v7g8 drosophila
16	241.5	6.1	9270	5	Q8mld9 drosophila

17	231	5.9	3111	5	Q9vh10
18	223.5	5.7	2083	5	Q9n435
19	223.5	5.7	10578	5	Q8isf5
20	223.5	5.7	18519	5	Q8isf6
21	223.5	5.7	18534	5	Q8isf7
22	223	5.7	2465	5	Q8iq71
23	222	5.7	1313	10	Q9xip6
24	222	5.7	1420	5	Q8inm3
25	222	5.7	1430	5	Q9vgz9
26	221	5.6	4723	5	Q8muJ8
27	220	5.6	811	4	Q9p2i3
28	218.5	5.6	200	10	Q04193
29	218	5.6	189	10	P93356
30	217	5.5	214	10	Q9SP34
31	216.5	5.5	1135	6	Q8hzw3
32	216	5.5	990	13	Q91803
33	216	5.5	1253	5	Q8sxm6
34	216	5.5	1298	5	Q9veu8
35	215.5	5.5	873	10	Q8GUP3
36	215.5	5.5	2139	5	Q07569
37	215	5.5	188	10	Q9SP33
38	215	5.5	985	5	Q9UOS5
39	215	5.5	1705	5	Q9UOS7
40	214.5	5.5	2910	10	Q9FND5
41	214	5.4	188	10	Q9XR99
42	214	5.4	1226	4	Q9NZB6
43	213	5.4	878	13	Q8AVW4
44	213	5.4	6632	5	O01761
45	213	5.4	6632	5	Q17362

ALIGNMENTS

RESULT 1
Q8K2H0 PRELIMINARY; PRT; 753 AA.
AC Q8K2H0;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC IONS.
DR EMBL; BC031490; AAH31490.1; -.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
DR PROSITE; PS00023; LIM DOMAIN 2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 753 AA; 84059 MW; 79F98B47C100CF22 CRC64;

Query Match 75.6%; Score 2970; DB 11; Length 753;
Best Local Similarity 75.9%; Pred. No. 5.1e-168;
Matches 577; Conservative 60; Mismatches 115; Indels 8; Gaps 3;

QY	1	MESSPFRQWTSLSLRVTA	KLSLVNKNKSAIV	EIFSKYKAAAEETNMEK	GRNTENL 60
DB	1	MESTFPNRQWTSLSLRVTA	KLSLVNKNKSAIV	EIFSKYKAAAEETNMEK	KQNPESL 60
QY	61	SOHFRTGLTTLVKKWENPGL	GAESHTSLRNS	STFIRHRADHP	PAEVTSHAASAKADQ 120
DB	61	POHFRGTLVLKKWENPVGA	EPTDLPNS	SEGHTADYPPAEVT	DKFAPGVRAD 120

```

QY 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKHSTESKMKMENCIGESRHEVEKSEISEN 180
DB 121 EEHTQPKRPFGRREAVIQSRYPSENHDPKQATSEQKMKENCIGDSRHEAEKPEITSEN 180
QY 181 TDASGKIKYVNPVNLRLKMWFEKPEPTQTKILRAQSASGSKISSENSYSLDDLEIGPQQ 240
DB 181 TETSGKIEKYVNPVNLRLKMWFEKPEHTQTKILRAQSASGSKISSENSYSLDDLEIGAGH 240
QY 241 LSSSTFDEKKNESRRNLPLRLSETS:IKDRMAKYAAVSKQSSSTNYTNELKASGGI 300
DB 241 LSSAFNFKKNESRRNLPLRLSETS:IKDRMAKYAAVSKQSSSTNYTNELKASGGI 300
QY 301 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 360
DB 301 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 360
QY 359 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 358
DB 359 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 358
QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHPRHDKDLWASKNENEELERPAQ 480
DB 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHPRHDKDLWASKNENEELERPAQ 480
QY 481 LANARETPHSQGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKCLRIAWPPPELGG 540
DB 481 LANARETPHSQGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKCLRIAWPPPELGG 540
QY 541 SGSALEBEGIKVSKPWPEDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQST 600
DB 541 SGSALEBEGIKVSKPWPEDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQST 600
QY 599 SIKSPKASSPLRKMGWSESESEFEGGATWERTKOTENARPSGKENVKSRWQ----- 653
DB 599 SIKSPKASSPLRKMGWSESESEFEGGATWERTKOTENARPSGKENVKSRWQ----- 653
QY 660 GETGKRSKEGHSLEMENENLVENGADSDSDNSFLKQSQPQPKSLNWSFVDNTFAEF 719
DB 660 GETGKRSKEGHSLEMENENLVENGADSDSDNSFLKQSQPQPKSLNWSFVDNTFAEF 719
QY 720 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 759
DB 720 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 759
QY 714 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 753
DB 714 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 753

```

RESULT 2

```

Q8C7S2 ID Q8C7S2 PRELIMINARY; PRT; 753 AA.
AC Q8C7S2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL; AK049350; BAC33699.1;
SQ SEQUENCE 753 AA; 84089 MW; 39571A17DF21F2C0 CRC64;

```

Query Match

```

Best Local Similarity 75.2%; Score 2955; DB 11; Length 753;
Matches 575; Conservative 60; Mismatches 117; Indels 8; Gaps 3;

```

```

QY 1 MESSPPNRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMKGRNTENL 60
DB 1 MESTPPNRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMKGRNTENL 60
QY 61 SQHFRKGTITVLKKWENPCLGAESHTDSLNRSSSTEIRHRADHPPAEVTSHAASGAKAQ 120
DB 61 POHFRKGTITVLKKWENPCLGAESHTDSLNRSSSTEIRHRADHPPAEVTSHAASGAKAQ 120
QY 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKHSTESKMKMENCIGESRHEVEKSEISEN 180
DB 121 EEHTQPKRPFGRREAVIQSRYPSENHDPKQATSEQKMKENCIGDSRHEAEKPEITSEN 180
QY 181 TDASGKIKYVNPVNLRLKMWFEKPEPTQTKILRAQSASGSKISSENSYSLDDLEIGPQQ 240
DB 181 TETSGKIEKYVNPVNLRLKMWFEKPEHTQTKILRAQSASGSKISSENSYSLDDLEIGAGH 240
QY 241 LSSSTFDEKKNESRRNLPLRLSETS:IKDRMAKYAAVSKQSSSTNYTNELKASGGI 300
DB 241 LSSAFNFKKNESRRNLPLRLSETS:IKDRMAKYAAVSKQSSSTNYTNELKASGGI 300
QY 301 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 360
DB 301 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 360
QY 359 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 358
DB 359 HKMEQENVPPEPEVCITHOEGEKISANENSILAVRSTPAEDDSRDSDQVSEVQVHPKP 358
QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHPRHDKDLWASKNENEELERPAQ 480
DB 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHPRHDKDLWASKNENEELERPAQ 480
QY 481 LANARETPHSQGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKCLRIAWPPPELGG 540
DB 481 LANARETPHSQGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKCLRIAWPPPELGG 540
QY 541 SGSALEBEGIKVSKPWPEDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQST 600
DB 541 SGSALEBEGIKVSKPWPEDEISKEPVEDVDLKLRRSSSLKERSRPTVAASFQST 600
QY 599 SIKSPKASSPLRKMGWSESESEFEGGATWERTKOTENARPSGKENVKSRWQ----- 653
DB 599 SIKSPKASSPLRKMGWSESESEFEGGATWERTKOTENARPSGKENVKSRWQ----- 653
QY 660 GETGKRSKEGHSLEMENENLVENGADSDSDNSFLKQSQPQPKSLNWSFVDNTFAEF 719
DB 660 GETGKRSKEGHSLEMENENLVENGADSDSDNSFLKQSQPQPKSLNWSFVDNTFAEF 719
QY 720 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 759
DB 720 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 759
QY 714 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 753
DB 714 TTQNKSDQVGLWEGEVVVELSVVEEQIKRNYDEDEDEE 753

```

RESULT 3

```

Q8CD09 ID Q8CD09 PRELIMINARY; PRT; 593 AA.
AC Q8CD09;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```


QY 161 MENCLOESRHEVEKSEISENTDASGKIEKYNVPLNLKMMFEKGEPTQTILRAQSRAS 220
Db 1 MENCLOESRHEAEKPEISENTTETSGKIEKYNVPLNLKMMFEKGEHNTQSLWTSRAG 60
QY 221 GRKISENSYSLDLEIGPGGLSSSTDSKNEKESRRNLELPRISSETGIKDRMAKYQAAVSK 280
Db 61 GRKISENSYSLDLEIGPGGLSSSTDSKNEKESRRNLELPRISSETGIKDRMAKYQAAVSK 120
QY 281 QSSSTNYTNELKASGGEIKIKHVEQKENVPPGPEVCITHQEGEKISANENSIAVASTPAE 340
Db 121 QSSPASVYNELKTS--ESKTHKWEQKENVPPGPEACSVHOGESKVSTTENSIALVSVAE 178
QY 341 DSDRDQVSKVEQVPHKPLSPDSRASSLSSESPPKAMKKFOAPARETCVSCQKTVYPM 400
Db 179 DDCNSQVSKSEAQQPHKPLSPDARTPLSPSSPKTAKKFOAPAKESCVCQKTVYPM 238
QY 401 ERLANQOVPHISCPCSYCNKLSLGTVAASLHGRYCKPHFNOLPKSGNVDGFGHRP 460
Db 239 ERLANQOVPHISCPCSYCNKLSLGTVAASLHGRYCKPHFNOLPKSGNVDGFGHRP 298
QY 461 HKDLWASKNEBILERRPAQANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKED 520
Db 299 HKDLWASKNEBILERRPAQANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKED 358
QY 521 KPAETKKLRJAWPPTELGSSGALBEGIKMSKPKWPPDEISKPEVPDVLKLR 580
Db 359 KPAETKKLRJAWPPTELGSSGALBEGIKMSKPKWPPDEISKPEVPDVLKLR 418
QY 581 SSKLKERSRFTVAASFOSTSVKSPKTVPPPIKKGWMSQSESVGG-RVAERKOVENA 639
Db 419 SSKLKERSRFTVAASFOSTSVKSPKTVPPPIKKGWMSQSESVGG-RVAERKOVENA 478
QY 640 KASKKNGNCKTWTQNKESGEGTGSKEGHSLEMEENLVENGADSDDDNSFLKQSP 699
Db 479 RPGEKENVCKSRQ-----GEEVPSKDRSSFELESEFMENGANIEDDHHVHAQQSP 533
QY 700 QBPKLWNSFVNTFABEFTTQNKSQDVVELWEGEVKLSVEEQIKNRYDEDEDEE 759
Db 534 LEPEAPGWSGFVDTTAAKEFTTQNKSDQVGFWEGEVWRELSVEEQIKNRYDEDEDEE 593
RESULT 6
Q9DEY8 PRELIMINARY; PRT; 629 AA.
AC Q9DEY8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytoskeleton-associated LIM domain protein.
GN EPLIN.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=1100452; PubMed=11179679;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
RT and comparison of mammalian and zebrafish EPLIN";
RL Gene 262.155-160(2001).
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR ENBL; AF307846; AAG31149.1; -.
DR HSP; P32965; 1CTL.
DR ZFIN; ZDB-GENE-001120-1; eplin.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.

DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 629 AA; 70037 MW; C42341B024818C03 CRC64;
Query Match 28.7%; Score 1128; DB 13; Length 629;
Best Local Similarity 36.9%; Pred. No. 6.5e-59;
Matches 287; Conservative 110; Mismatches 212; Indels 168; Gaps 28;
QY 1 MESSFPNRRWTSLSLRVTAKELSLVN-KNKSAIVEIFSKYQAAEETNMKKRSNTEN 59
Db 1 MVSFFRRGQWASQSLRVTAKEISLVGRGKNTAIAERFSKYQAABETSIDKKKS-PEK 59
QY 60 LSOHFRKQTLVLKKKNENPGLGASHTDSLNRNSTIRHRADHPPEAVTSHAASGAKAD 119
Db 60 STPLGRNLNLVLKQLEHP---AETPT-----SPEPKAHLQNLH 96
QY 120 QEOIHPHSRLRSPEALVQGRYPHIKDGEDLKHSTESKMCNCLGESRHEVEKSEISE 179
Db 97 --QASAVKIPLESTDVQJIEG-----TD-----QQCLSDSDQPMK---RT 132
QY 180 NTDASGKIEKYNVPLNLKMMFEKGEPTQTILRAQSRASGRKISENSYSLDLEIGPG 239
Db 133 QRDVETLLEKPTVPLNSLKMMFEKGETLHNSVRE-----PG 169
QY 240 QLSSSTPDSKNEKESRRNLE--LPRLSSETSIKDRMAKYQAAVSKSSSTNYTNELKASGE 297
Db 170 TTGSGSDNMPEPTGESLECGVKMLDSTPLRDRVAMYQAAVTKLDFFSSPNSE--AADSE 227
QY 298 IKIHMEQKENVPP-GPEVCITHQEGEKISANENSIAVASTPAEDDSDSOVKSEVQPV 356
Db 228 ARAHSGKQKENVPPVSADV-----PESNTIKSP-----256
QY 357 HPKPLSPDSRASSLS-ESSPEKAMKKFOAPARETCVEQKTVYPMERLAAVQVPHISCF 415
Db 257 -----APDRNGSVLSPEQKQPLVMPFLPVRETCTVCLTVYPLEKLVANQQIYHNTCF 311
QY 416 RGSYCNKLSLGTVAASLHGRYCKPHFNOLPKSGNVDGFGHRPHKOLWASKNE--NEE 473
Db 312 RCAYCNTKSLVNVASLHNNYCKPHYCQLFKAGNYDEGFGHRPHKELWEGRPEGVDQ 371
QY 474 ILERPAQANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRJAWP 533
Db 372 VKLSP-----QET--SLTVEESPLVKNVLAATLETQTATSERVEKPLETGKLSWP 423
QY 534 PTELGGSSGSL--EEGKMSKPKWPPDEISKPEVPDVLDDKLRRSSSLKERSRP 590
Db 424 PQSEGESATHVTGSGIKPIRKWPPGDTVSSNV--DUESDLPKURRSVLSKERSKP 481
QY 591 FTVAAFSQTSVKSPKTV--SPPIKKGWMSQSESVGGVRAERKQVENAKAKKNGNV 648
Db 482 FCI---FDSAPVAQPKRCQSPSNEKPDSEEMSPVS-----STDTTISSED---526
QY 649 GKTWTQNKESGEGTGSKEGHSLEMEENLVENGADSDDDNSFLKQSPQE--PKSLN 706
Db 527 --MTEHQSEEDQDKTKEDEQMEHBEK-----VDVQEEELSLKCSPTDNPPLSPE 577
QY 707 WSSPVDNTFABEFTTQNKSDQVVELWEGEVVKE-----LSVEEQIKNRYDEDEDEE 759
Db 578 SESGLD-----PEENQASQDVGFWDGEEAEDTADVTVEDLTKRNRHYDDEDED 627
RESULT 7
Q8HXJ9 PRELIMINARY; PRT; 951 AA.
AC Q8HXJ9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.


```

OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RX MEDLINE=21458551; PubMed=11574149;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
RT chromosomes.";
RL Gene 275:31-37(2001).
DR EMBL; AB097518; BAC41743.1; -.
SQ Hypothetical protein.
SQ SEQUENCE 951 AA; 107682 MW; 42A3F8DECO5274E2 CRC64;

Query Match 16.3%; Score 641; DB 6; Length 951;
Best Local Similarity 27.7%; Pred. No. 8.3e-30;
Matches 212; Conservative 102; Mismatches 221; Indels 230; Gaps 28;

QY 65 RKGTITLVKKWENPG-----LGAESHTDSLNRNSSTEIRHDPHPAEVTSAAAGAKA 118
DB 19 QKGSUNLLRQKESCGYORSECPYRDSRCTIIQPOESKLL-----EPSEVVS----- 66

QY 119 DDEEQIHPRS-----RLRSPEALVQGRYPHIKDGEDLKDHSSTESKXNCLGESRHE 171
DB 67 -APEPLDPTSLPCSGELTSSKPE-----GKDSVDKSNNTREY-----GRPE 107

QY 172 VKSEISENTDASGKTEKYNVPLNRLKMFKEGEBTQTKILRAQSRASGRKIS-ENSYS 230
DB 108 VLK---EDSLSSRRRIERFESIALDELRSVFAPKSGN-----KSAHGGKVEIERSLUC 158

QY 231 LDDLEIGPG-QLSSTFFDSE-----KNESSRNLEUPR 261
DB 159 STAFKSHPGSLEDVSKDSDEKGEETSCDKMPSGSHSHIFPAIAGPKNPKSFGFEDSAA 218

QY 262 LSE-----TSIKDRMAKYQAASVK---QSSSTNYTNELK---ASGGEIKIHMEQKE 307
DB 219 LGEVSDHLHVSLKERMARYQAASVGRGCRSFSANMWESEMCVTPGGLAKVKQFDE 278

QY 308 NVPPGPEVCIT-----HGEKISANENSILAVRSTPAEDDRDSQVKSEVQO- 354
DB 279 -----ITSSRNTFAQYQYQHQRSEQAIHSSQVGTSSKSOELARNEQEGSKVQKI 329

QY 355 -----PVPKPLPSDSR----- 366
DB 330 DVHGTGEMVSHLEKHTKEINQASQFHQYQVETVIDTPEDEEIPKYSTKLLKEQFEKSAQEK 389

QY 367 --ASSLSSESPKAMKFFQAPARETCVECKQTVYPMERLLANQOVFHSICRSCSYCNKL 424
DB 390 ILYSKEMTTPAKQIKLLQDKEICILCKQTVYPNECLVADKQNFHAKCFCHHCNSKL 449

QY 425 SLGTVYASLHGRYCKPHFNQLPKSGNYDEGFGHPRPHKDLWASKNENEEI-----LERPAQ 480
DB 450 SLGNVYASLHGRYCKPHFNQLPKSGNYDEGFGHPRPHKDLWASKNENEEI-----LERPAQ 509

QY 481 LANARETHPSGVEADIAKVGLVAASMEAKASSQOEKE-DKPAETKXLRIAWPPPTLGL 539
DB 510 CKNTAENTLVPGDRN-----EHLDAQSGEGRNDRKLGKGLKVIWPPSKVIP 559

QY 540 SSGSALEEGIKMSKPKWPPEDEISKPEVPEVDYLDLKKLRSSSLKERSR----- 589
DB 560 KKTFFPEELRWSKPKWPP--EMTTPLSPE-----FKSESLEDVTRTPENKGGQED 608

QY 590 --PFTVAASQSTSVKSPKTVSPPIRKGMWSMEQSEESVGGVRAERKOVENAKASKNGN 647
DB 609 HLPF-LQPLQSTHVC-----QKEDVIG--IKEMQKHEVRKDEKB-- 646

QY 648 VKGTTWQNKESGKETGKSKSGHSELEMNENIENVGADSDDEDNS 692
DB 648 VKGTTWQNKESGKETGKSKSGHSELEMNENIENVGADSDDEDNS 692

```

```

DB 647 -GRKNVQDRLSEABDTKSNRSEKSEMDLNDNHNHVVQSAEKEKEKNEKT 690

RESULT 8
Q8N720
AC Q8N720; PRELIMINARY; PRT; 519 AA.
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
OS Hypothetical protein FLJ40200.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sakine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahari K., Masuhara Y., Nagai K., Isegai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; AK097519; BAC05086.1; -.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
DR PROSITE; PS00023; LIM DOMAIN 2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 519 AA; 58797 MW; 69DE84BB60E9106 CRC64;

Query Match 13.5%; Score 529; DB 4; Length 519;
Best Local Similarity 29.9%; Pred. No. 1.7e-23;
Matches 160; Conservative 66; Mismatches 170; Indels 140; Gaps 19;

QY 65 RKGTITLVKKWENPGGLGAESHTDSLNRNSSTEIRHDPHPAEVTSAAAGAKADQEQI 124
DB 5 QKGSUNLLRQKWE-----SCDYQSRSECHPR---DSHCTI-FQPOESKLL 44

QY 125 HPSRLSPPEALVQGRYPHIKDGEDLKDHSSTESK---KXNCLGSRHEVEKSEISEN 180
DB 45 APEGEVVSAPQSLDPTSLPY-STGEEMWSKPEEKSDVSKSNNTREYGRPEVLK---EDS 100

QY 181 TDASGKIEKYNVPLNRLKMFEE-----KGEPTQTKIL 212
DB 101 LSSRRRIERFESIALDELRSVFAPKSGNKPAYGKVEIERSLCSFAPKSHPGSQLK-- 158

QY 213 RAQSRASGKISSENSYSLDLEIGPQGLSSSTFDSSEKNS-----RNLELPRLE- 264
DB 159 --DSVKDSKKGKETSFDKVPSPGSHRIFEATAGPNKPSGFAEDSAAAGGVSDLHEV 216

QY 265 TSIKDRMAKYQAASVK---QSSSTNYTNELK---ASGGEIKIHMEQKE----- 307
DB 217 VSLKERVARYQAASVGRGCRSFSANMWESEMCVTPGGLAKVKKQFDEITSSRNTFAQY 276

QY 308 -----NVPPGPEVCITHQGEKISANENSILAVRSTP-----AE 340
DB 277 QYQHQRNSEQAIHSSQVGTSSQEMARNEQEGSKV---QKIDVHGTGEMVSHLEKHT 332

QY 341 DDSRDSQVKSEVQOVPHPK-----LSPDSFASLSE-----SSPPKAMK 380
DB 333 EVQASQFHQYQVETVIDTPEDEEIPKYSTKLLKEQFEKSAQEKILYSKEMTTPAKQIK 392

QY 381 KFOAPARETCVECKQTVYPMERLLANQOVFHSICRSCSYCNKLISLCTYASLHGRYCKP 440
DB 393 KLLQDKEICILCKQTVYPNECLVADKQNFHAKCFCHHCNSKLSLGNVYASLHGRYCKP 452

```

QY 441 HFNLKSKGNVDEGFGHPRPHKDLWASKNENBEI-----LERRAQLANARETTPHSPG 492
DB 453 HFQLFKSKGNVDEGFGHPRPHKDLWASKNENBEI-----LERRAQLANARETTPHSPG 508

RESULT 9
Q8BGB5 PRELIMINARY; PRT; 128 AA.
AC Q8BGB5
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical LIM domain/LIM domain profile/cytochrome c family
DE heme-binding site containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body, Hippocampus, and Olfactory brain;
RX MEDLINE=22354683; PubMed=12456851;
RA The PANTOM Consortium
RT "Analysis of the mouse transcriptome based on functional annotation of
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT Nature 420:563-573(2002)."
RL Nature 420:563-573(2002).
DR EMBL; AK012581; BAC25371.1; -
DR EMBL; AK032430; BAC27866.1; -
DR EMBL; AK049809; BAC33928.1; -
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 14237 MW; ADF9161771331D13 CRC64;

Query Match 9.2%; Score 363; DB 11; Length 128;
Best Local Similarity 58.7%; Pred. No. 2e-14;
Matches 64; Conservative 16; Mismatches 27; Indels 2; Gaps 1;

QY 362 SPDSRASSLSSESPPKAMKRF--QAPARETCVECKTVPMERLLANQVPHISCFRCYS 419
DB 11 TPSHEAKGSSGSSVQSRKSFSLRAQVKETCAACQKTVPMERLVADKLIFHNSCFCKH 70

QY 420 CNKLSLGTYSALHGRYCKPHFNQPKSKGNVDEGFGHPRPHKDLWASK 468
DB 71 CHTKLSLGSVAAMGEFYCRPHFQOLFKSKGNVDEGFGHPRPHKDLWASK 119

RESULT 10
Q9BT23 PRELIMINARY; PRT; 127 AA.
AC Q9BT23
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein FJ34982.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., Ono Y.,
RA Horita T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; BC004400; AA04400.1; -
DR EMBL; AK092301; BAC03855.1; -
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM DOMAIN_1; 1.
DR PROSITE; PS00478; LIM DOMAIN_2; 1.
DR PROSITE; PS00478; LIM DOMAIN_3; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14070 MW; B63174FCF0486956 CRC64;

Query Match 9.1%; Score 357; DB 4; Length 127;
Best Local Similarity 50.7%; Pred. No. 4.4e-14;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDDSRDSQVSKSEVQVPHKPLSPDSRASSLSSESPPKAMKKFQAPARETCVE 392
DB 7 AAQATPSHDAKGG--SSTVQ-----RSKSFS-----LRAQVKETCAA 42

QY 393 CQKTVPMERLLANQVPHISCFRCYSYCNKLSLGTYSALHGRYCKPHFNQPKSKGN 452
DB 43 CQKTVPMERLVADKLIFHNSCFCKHCKHTKLSLGSVAALHGEFYCKPHFQOLFKSKGN 102

QY 453 DEFGHPRPHKDLWASK 468
DB 103 DEFGHPRPHKDLWASK 118

RESULT 11
Q96S91 PRELIMINARY; PRT; 127 AA.
AC Q96S91
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Zhang W., Zhang M., Wan T., Cao X.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; AY037154; AAK67634.1; -
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM DOMAIN_1; 1.
DR PROSITE; PS00478; LIM DOMAIN_2; 1.
DR PROSITE; PS00478; LIM DOMAIN_3; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14142 MW; B63014FEF0486954 CRC64;

Query Match 9.1%; Score 357; DB 4; Length 127;
Best Local Similarity 50.7%; Pred. No. 4.4e-14;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDDSRDSQVSKSEVQVPHKPLSPDSRASSLSSESPPKAMKKFQAPARETCVE 392
DB 7 AAQATPSHDAKGG--SSTVQ-----RSKSFS-----LRAQVKETCAA 42

```
QY 393 CQKTVYPMERLLANQOVFHISCFRCSYNNKLSLGTYSALHGRYICKPHFNOLFKSKGNY 452
Db 43 CQKTVYPMERLVADKULIFHNSCFCKCHTKLSLGSVAALHGFYCKPHFQQLFKSKGNY 102
QY 453 DEQFGHPRPKDLWASK 468
Db 103 DEQFGHPRPKDLWASK 118

RESULT 12
Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T198.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RL MEDLINE=20363099; PubMed=10907853;
RX "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1;
SQ SEQUENCE 2081 AA; 232851 W; D3603ELP8SEFF29 CRC64;

Query Match
Best Local Similarity 20.6%; Pred. No. 1.1e-06;
Matches 178; Conservative 143; Mismatches 317; Indels 226; Gaps 37;

QY 13 SLSURVAKELSLVNKKSSAIVFIPKYQAAETNWEKKRSNTENLSQHFRKGLTULV 72
Db 450 NLENKVGNEEL-----KGNASVEAKTNBESSKEEKSORSNEVYNNKTTKG----- 498
QY 73 KKKWENPGLAESHTDSLRSSTE-----IHRAD-----HPPAEVTSHAASGA 116
Db 499 -----ENVNIQGESIGDSTKNSLENKEDVKPKVDANESDGNSTKERHQBQVNNGVSTED 554
QY 117 K-----ADQEQIHPRSLRSPPEALVQGRYPHIKD-GEIDLKDHSTESKWMENCLGESR 169
Db 555 KNLDNICADEQKK-----NDKSVEVTNDGDHTKREKRETQNGESVKNENL-----E 603
QY 170 HEVKSSEISENTDASGKIEKYNVPLNKLKMMFEKPTQ-----TKILRAQSRASG 221
Db 604 NKEDKELKDDSEYGA-----TNNETSLEEKREQTQKGDHNSINSKIVDNKGGNADS 656
QY 222 RKISE-----NSVSLDD-----LEIGPQLSSSTFSEKWNESRRNLLEPL-----S 263
Db 657 NKEKEVHVGSDTNDNNMESKEDTKSEVEVKNDGSEKGEKGNKNSMEDKKLENKES 716
QY 264 ETSIKDRM-----AKYQAAVSKOSSSTNYTNELK-----A 293
Db 717 QTDSDKDDKSVDDKQEEAQIYGEISKDDKSKVEAKGKKESKENKTKT-TNENRVNKEENV 775
QY 294 SGGGIKIHMEQKENVPPGPEVCITHOGEKISANENSLAVRSPPABDSDRDSQVSKSEVQ 353
Db 776 QGNKKESEKVBKESKSKDAKSVETKDKNKLSTENRDEAKERSGDNKEDKE-ESKDY 834
QY 354 QPVHPKPLSP-----DSRASSLSSESPPKAMKFOAPA-RETVCQKTVYPMERLLANQ 408
Db 912 QPVHPKPLSP-----DSRASSLSSESPPKAMKFOAPA-RETVCQKTVYPMERLLANQ 408
```

```
Db 835 OSVEAKEKXNNGVDNTVGNKEDSKDLKDDRSVVEVKANKESMMKKKREEVORNDKSSTKE 894
QY 409 VFHISCFRCSYNNKLSLGTYSALHGRYICKPHFNOLFKSKGNYDEGFGHPRPKDL- 464
Db 895 V-----RDFANNWDIDVQKSGSEVK---YKDE-----KKEGKEE-----NKDTI 933
QY 465 WASKNENEIEILRPAQLANARETPHSPGVEDAPIAKYGVLAASWEAKASSQOE----- 517
Db 934 NTSSKQKQKDKKKKKESKN-----SNMKKKEEDKKEYVNNEL 971
QY 518 ---KEDKPAETKKLRIAMPPTTELSSGSGSALBEGIKMSKPKWPPPEDEISK-----PEVP 568
Db 972 KQEDNKKETTK-----SENSKLKENKDKKEKSEDSASKNREKKEYEKK 1019
QY 569 EVDVLDLKLRRSSSLKRSRPFVTVAASFQSTSVKSPKTVPPPIKGMMSFSQSESVGG 628
Db 1020 SXTKEEAKKEKKKSQDKKREE-----KDSEERKSKKEEESRDLK-----AKKKEE- 1065
QY 629 RVAERKQVENAKASKNGNVGKTTWQNKESKGETGKRSKEGHSLEMEENLVENGADSD- 687
Db 1065 ETKEKKESENHKKKEDYK-KEHEDNKSMMKKEEDKKEKKHE-ESKRRKKEEDKKDMEK 1122
QY 688 -EDNNSFLKQSQPQBPFGKSLNWSFYDNTFAEFTTQNG-----KSDVLEWMEGE 735
Db 1123 LEDQNSNKKKEDKNEKKKSHQVKKESDKKEKKEBESKSETKIESSKSQKNEVDKKE 1182
QY 736 VVKELSVBEQIKRNYDEDEDEE 759
Db 1183 --KXSKDQKQKKEKEMKESEBK 1204

RESULT 13
Q9W596 PRELIMINARY; PRT; 5412 AA.
AC Q9W596;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG3064 protein.
GN FUTSCH OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Barnko C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gortell J.H., Gu Z., Guan P., Harris M.,
RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jalaali M., Kaulush F., Karpen G.H., Ke Z., Kenson J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
RT "Sequencing of *Drosophila melanogaster* genome,"
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of *Drosophila melanogaster* genome,"
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003420; AAF45622.3;
DR FlyBase; FBN0015390; futsch.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR000531; TonB_boxC.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
SQ SEQUENCE 5412 AA; 584513 MW; 7881CAAC8749FAFF CRC64;

Query Match 6.4%; Score 251; DB 5; Length 5412;
Best Local Similarity 20.7%; Pred. No. 1e-05;
Matches 199; Conservative 138; Mismatches 359; Indels 264; Gaps 37;

QY 22 ELSLVNKNKSSAIVFISFYQKAAETNMKRSNTNLSOHFRKGTILVLKKWNPGL 81
DB 954 EDSIVQESSMTKEEIQHQDSQSESEKKKSAEEIEIAIAKVAEAKRALE---- 1009
QY 82 CAESHTDSLRSSTIRHRAHPHVAETSHAAG---AKAQEEQI--HPRSRLRSP-PE 135
DB 1010 GASARQD---ESELDEVEPQSKIAEVQDIATATAKDIKRSIEQLAKPAEELSSTPE 1066
QY 136 ALVQGRYPHIKQGE-----DLKHSTESK-----X 160
DB 1067 EKLSKKTSQDDQIGAPVDLPVNLQESLPSEKFSATIESGATTAPLPLDERIPLDQI 1126

QY 161 MENCIGSRHEVEKSEISE-----NTDASGKI 187
DB 1127 KEDLVIEEKYKKEETKAEIAVVAIVQTLPEAPLAIDTILASATKDAPKANAAALGEL 1186
QY 188 EKYNVPLNRLKNMFKGEPQTQKILR-----AQSRSASGRK 223
DB 1187 PDSGERVLPKMTF-----AQQLLRDVIKTPDEVADLPVHEEADLGLYKDSQDAGAKS 1242
QY 224 ISENSYSL-----DLEIGFGQLSSSTFDBSEKNESRNLEPLRLSETSIKDRVAKYQA 276
DB 1243 ISHKEESAKEKETDDEKENKVGTE---LGDEPNKVDISHVLLKESVQVEAKVVVIET 1299
QY 277 AVYSQ-----SSSTNYTHLKAASGEIKLHKMEQENVPVPGPEVCITHQEGEKISANENS 331
DB 1300 TYEKQEEIVEATTITQENQEDLMQVKDKEEHEQKTESG---ITKEAKKSA----- 1351
QY 332 LAVRSTPAS-----DSDRSDOVKSEVQVQHPKPLSPDGRASSLSSESP-----K 377
DB 1352 ----STPESEKTSIDTSDDELPQAADPTTPPKSAKREDTSGSTIESPTIEEAIEVEVQ 1407
QY 378 AMKKFO--APARETCVECOCTVYPMERLNAQQVFHISCFRCSCYNNKLSLGTYSALHGR 435
DB 1408 AKQEAQKVPAPPEAAIKTEKSP-----LASKETSRPESATGSV---KEDTQTKSKXSP 1458
QY 436 IYCKPHFNOLFKSKNYDEFGHRP-----HKDLWASKNENEIEILERP 479
DB 1459 VPSRPE-SEAKDKGPPFASGEASRPESVAESVKAESRRSRESIAKTHKDESLDKAK 1517
QY 480 QLANARETHSPGVEDAPIAKVGLAASMEAKASSQOEKEDK-----AETKKL 528
DB 1518 EQESRRRESIAESIKPESGIDEXASLASKEASRPESVTDKSKPESRESIAESLKAESTKD 1577
QY 529 RIAMPPTTELGSAGSALE-----EGIKWMSKPK-----WPP-----E 559
DB 1578 EKSAAPSKASRPGSVVSVKDETEKSPRESIAESAKPIEPREVSRPESVIDGIK 1637
QY 560 DEISKPEVPEDVDLKLRRSSSL-----KERSRPTVAASQTSQSVKSPKTV 608
DB 1638 DESAKPESRRDPSLAKESASRPESVLESVKDEPIKSTEKSRRESVAESFKADSTKDEK-- 1695
QY 609 SPIRKGNMSQSESVGGRV--ABRKQVENAKSKNGNVTGKTTWQNKESKGETGKRS 666
DB 1696 SPLTSKDISRPESAVENVMVDAVGSASRPESVTSR---DVSRPESVAESKDDTDKPE 1752
QY 667 KECHSLEMENL--VENGADSDDEDNSFL-----KQSQPE-----PKSLNW 707
DB 1753 SVVESVIPASDVVEIEKGA-ADKEKGVFVSLGIRPDSFSEVISRPGVPVSVKPSRRE 1811
QY 708 SS--FVDNTFAEEFTTQNGKSDQDELW--EGEVVK-----ELSVEEQIKENRYDEDEEE 759
DB 1812 SSTEIVLPCHAEKSPESKVECLKDESEVLKSGSTRRESVAESDKSSQPPKTSRPE 1871

RESULT 14
QY 076891 PRELIMINARY; PRT; 5327 AA.
AC 076891;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE EG:49E4.1 protein.
GN FUTSCH OR EG:49E4.1 OR CG3064
OS *Drosophila melanogaster* (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.,
RT "Sequencing the distal X chromosome of *Drosophila melanogaster*,"
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031128; CAA20006.1; -
DR FlyBase; FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFSE23A118FF38A CRC64;

Query Match 6.3%; Score 246; DB 5; Length 5327;
Best Local Similarity 22.3%; Pred. No. 1.9e-05;
Matches 192; Conservative 124; Mismatches 330; Indels 214; Gaps 38;

QY 77 ENPGLCAESHTSLNSSTFIRHRAHPAETVSHAAGAKAQDEQIHPRSLRSLSPSEA 136
DB 2600 ERPSPAESAADAA--ESVEKSKDASRPSPSVVSTKADSTKGDISPS--PSVLEGPDKD 2655

QY 137 LVQ----GRYPHIK---DGEDLKDHS-----TESKMKMNCLOESRHE-----VE----- 173
DB 2656 VEKSKESRRPPSVASITGDSKDVSRPASVSVESVKDEHDKAESRRESIAKVESVIDEAG 2715

QY 174 KSEISNTDASQIKBYNV---PLNRLKMMFKGPTQTQI-LRAQSRASAGRKISENSY 229
DB 2716 KGDSSSSQDSQDKBSTLASKEASRRRESVWSSKDDAEKSSRSPESVIAAGEPVPRESK 2775

QY 230 S-LDLEIG-PCQ-LSSSTFDSEKNE--SRRNLELPRLSETSIKDRMAKYQAA----- 277
DB 2776 SPLDSKOTSRPGSVSVESVTADEKSEQQRRESVSAESVKAADTKD--GKSOEASRPSVD 2833

QY 278 -----VSKQSSSTNYTNELKASGEIKIHKMEQKENVPPGPEVCITHQGEKISA 327
DB 2834 ELLKODDEKQESRRSITGSHKAMSTMGDESMPDKAD-KSEKSPSPESVSAIKHENTKD 2892

QY 328 NENSILAVSTPAEDDSRDSQVKEV-----QQVHPKPLS-----PDSRASLL 370
DB 2893 EESPGLSRA-----DSVAESIKSDITKGEKSLPSKEVSRPSVSGSIKDEKAESRRESV 2947

QY 371 SSSSPKAMKTKQA--PARETC-----VEQKTVYPMERLLANQ 408
DB 2948 AESVPESSKDATAPPSKEHRSRPSVGLSKDEGDKTTSRVSADSIKDEKSLLVQ 3007

QY 409 VFHISFCRCSYCNKLSTGYASLHGRYCKPH-----FNQLFK 447
DB 3008 -----ASRSESAESLKDAAAPSOETSRPESVTSVKDGKSPVAKESASRPASVAEN 3059

QY 448 SKGNVDEGHRP-----HKDLWASKHNEEIEILERPAAQLANARETPHSPGVE 494
DB 3060 AKDSADESKEQRPESILPQSKAGSIKDEKSLASKDEAKSRESRESVAEQFP-----LV 3115

QY 495 DAPIAKGVLAAMEAKASSQEKDKPAETKRLRIANPPPTLGGSSGSALEEGIKMSK- 553
DB 3116 SKEVSRPASVAESV--KDEAKSKESPLMSKE--ASRPASVAGSVKDEAKSKESRR 3170

QY 554 ----PKNP-PEDEISKP-----EVPEDVDLDLKKLRRSS-----SLKERSRPTVAAS 596
DB 3171 ESVAESPLPSKEARSPASVAESVKDEADKSESRRESGAEKSLASKESASRPASVAES 3230

QY 597 F-----OSTSVKSP---KTVSPPIRKGMMSSEQS-----EESVGRVRAERQV 636
DB 3231 KDEAKSKESRRRESVSAESKSLPSKEASRPTSVSAKSVKDEAKSKESRRSDSVAESKPL 3290

QY 637 ENAKASKMGNGVYKTTWQNKESGTGRSK--EGHSLMEENENLIVGASDDEDNFL 694
DB 3291 ASKEASRP-ASVAESQDEAKSKESRRRESVAESKSLPSKASRPASVAESIKDEAKS 3349

QY 695 KQSQOE-----PKSLNWSFVDNTFAEFTTQNKQSDQVLEWGEVYKE 739
DB 3350 KEESRRRESVAESKSLASKESRPTSVSAESVKDEAKSKESRRSDSVAESKSLASKESGRP 3409

QY 740 LSVVEQIKNRYDDEDEE 759
DB 3410 ASVAESV-----QDEAEK 3422

RESULT 15
Q9V7G8 PRELIMINARY; PRT; 7210 AA.
AC Q9V7G8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG18255 protein.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mourtoulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Schaefer F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]
SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman B., Carlson J.W., Celniker S.E.,
RA Clamp W., Drysdale C., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003908; AAF58087.2; -
DR FlyBase; Fgdn0013988; Strn-Mlck.
DR InterPro; IPR0031599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR Pfam; PF00047; Ig_13.
DR SMART; SM00409; Ig; 21.
DR SMART; SM00408; IGC2; 20.
DR PROSITE; PS50835; IG-LIKE; 20.
DR PROSITE; PS00290; IG-MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;

Query Match 6.1%; Score 241.5; DB 5; Length 7210;
Best Local Similarity 20.4%; Pred. No. 5.3e-05;
Matches 167; Conservative 155; Mismatches 357; Indels 139; Gaps 31;

QY 19 TAKELSLVNNKSSAIVEIFSKYQAAETNNKESNTENLSQHPKGTTLVLKKWEN 78
DB 3765 TOTDSAIDEKSAEYSEIVS--EKITDEKQESQKEEVKD-SEAKPKKAKVLEKSIIE 3821

QY 79 PGL--GAESHTDSL---RNSSTIHRADHPAETVTHAASGAKADQEQIHPRLRSP 133
DB 3822 EKLEDKKEKQTESAIDEKSAEYSEIVS--SEIVSEKITDEKQESQKEEVKSEAKP 3873

QY 134 PEALVQGRYPHIKDGEDLDKDHSTESKMNCLGESRHEVEKSEI-----S 178
DB 3874 KKAKVLEK-----KSIEEEKLEDKKQTESAIDEKSAEYSEIVSEKITDEKQESQME 3929

QY 179 ENTDSAGKIEKYNVPLNRLNMMEKGEPTQTILRAQSRASGRKISENSYSLDDLEIGP 238
DB 3930 EVKDSKPKKAKV-LEKKSIEEEKLENKKEK-----QTESA----IDEKSAEYSEI-- 3978

QY 239 GQLSSSTFTDSEKNESRRN-----LEPRLSETSIKORMAKY-QAAVSKOSS 283
DB 3979 --VSEKITDEKQESQKEEVKDSKPKKAKVLEKKSIEEEKLEDKKQTESAIDEKQ 4036

QY 284 -----STNYTNELKAGGEIKIHKMOKEKENVPPGPEVCITHOGEKISANENSLAVR 335
DB 4037 KAEVSEIVSENIITDE-KAQESQKK--EVKDSKPKKAKV-----EKKSTIEEEKLEDK 4087

QY 336 STPAEDDSRDSQ-VKSEVOQVHPKPLSPDSRASSL-----SESSPPKMKYFQAPARET 389
DB 4088 KETQTSADAIDEKSAEYSEIVSEKITDEKQESQKEEVKDSKPKKAKVLEKKSIEEE 4147

QY 390 CVECQKTVPMERLLANQQVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNOLFYSK 449
DB 4148 KLEDDKKEKQTESAIDEKSAEYSEIVSENIITDEKQESQKEEVKDSKPKKAKVLEK 4207

QY 450 GNYDEGFGHRPHKDLWASKNE-----NEETILERPAQLANARETPHSPG--VEDA 496
DB 4208 SIEEEKLEDKKQTESAIDEKSAEYSEIVSEKITDEKQESQKEEVKDSKPKKAK 4267

QY 497 PIKVGVLAAASMEAKASSQOEK--EDKPAETKKLRITAWPPPTTELGGSGSALBEGIKMSKP 554
DB 4268 VLEKKSIEEEKLEDKKETQDTSADAIDEKSAEYSEIVSEKITDEKQESQKEE-VKQSEA 4326

QY 555 KWPEDSISKPEVPEDVDLDLKKLRRS-SSLKERSRFTTVAASFQSTSVKSPKTVSPPIR 613
DB 4327 K-PKAKVLEKKSIEEEKLENKKEKQTESAIDEKSAEYSEIVSEKIT 4374

QY 614 KGWMSQSQSEESVG-----GRVAERKQVENAKASKKNGNKGKTTWQNKESKGETGKRS 666
DB 4375 DEKQESQKEEVKSEAKPKKAKVLEKKSIEEEKLEDK-----KEKQTESAIDEKS 4425

QY 667 KEGHSLMEENENLVENGADSDDEDNSFLKQOSQOEKSLNWSFVDNTFAEEFTTQ----- 722
DB 4426 QKAEVSEIVSEKITDEKQESQKEEVKDSKPKKAKVLEKKSIEEEKLEDKKETQDTS 4485

QY 723 -NQASQDVELWEGEVVKELSVEEQIKENRYDEDEEE 759
DB 4486 IDEKSAEYSEIVSEKITDEKQESQ-KEEVKQSE 4520

Search completed: January 6, 2004, 09:49:54
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:45:49 ; Search time 21 Seconds
(without alignments)
1529.235 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSFNRRQWTSLSRVTA.....LSVBEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS-COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	5.7	1312	4	US-09-345-882-29 Sequence 29, Appl
2	217	5.5	219	3	US-08-934-627B-6 Sequence 6, Appl
3	210.5	5.4	200	4	US-09-282-146-2 Sequence 2, Appl
4	205.5	5.2	208	3	US-08-934-627B-2 Sequence 2, Appl
5	193	4.9	2842	1	US-07-741-940-7 Sequence 7, Appl
6	193	4.9	2842	1	US-08-289-548A-7 Sequence 7, Appl
7	193	4.9	2842	1	US-08-452-654-7 Sequence 7, Appl
8	193	4.9	2842	4	US-08-449-731-7 Sequence 7, Appl
9	193	4.9	2843	1	US-07-741-940-2 Sequence 2, Appl
10	193	4.9	2843	1	US-08-289-548A-2 Sequence 2, Appl
11	193	4.9	2843	1	US-08-452-654-2 Sequence 2, Appl
12	193	4.9	2843	1	US-08-452-655B-2 Sequence 2, Appl
13	193	4.9	2843	1	US-08-452-655B-7 Sequence 7, Appl
14	193	4.9	2843	2	US-08-370-235A-2 Sequence 2, Appl
15	193	4.9	2843	3	US-08-450-582-2 Sequence 2, Appl
16	193	4.9	2843	3	US-08-450-582-7 Sequence 2, Appl
17	193	4.9	2843	4	US-08-449-731-2 Sequence 2, Appl
18	193	4.9	2843	2	US-08-821-353A-7 Sequence 7, Appl
19	193	4.9	2843	2	US-09-003-687A-7 Sequence 7, Appl
20	193	4.9	2843	4	US-09-136-605-7 Sequence 7, Appl
21	192	4.9	1780	1	US-08-769-309A-5 Sequence 5, Appl
22	192	4.9	1780	3	US-08-994-570-5 Sequence 5, Appl
23	190	4.8	1864	2	US-08-790-912-3 Sequence 3, Appl
24	190	4.8	2052	2	US-08-790-912-2 Sequence 2, Appl
25	187.5	4.8	3969	3	US-08-061-376-5 Sequence 5, Appl
26	184.5	4.7	1939	4	US-09-310-187A-1 Sequence 1, Appl
27	178.5	4.5	1848	3	US-08-296-791-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-345-882-29
; Sequence 29, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; CURRENT APPLICATION NUMBER: US/09345.882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 29
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 294..296
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 432..434
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 755..757
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 856..858
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 859..861
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 910..912
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1151..1153
; OTHER INFORMATION: potential
; FEATURE:

Sequence 6, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 32, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 11, Appl
Sequence 16, Appl
Sequence 2, Appl
Sequence 14, Appl

28 178.5 4.5 1848 5 PCT-US95-10661A-6
29 178 4.5 1269 4 US-09-645-456A-15
30 178 4.5 1269 4 US-09-425-324A-15
31 178 4.5 1269 4 US-09-645-791-15
32 177.5 4.5 1324 4 US-09-645-456A-13
33 177.5 4.5 1324 4 US-09-425-324A-13
34 177.5 4.5 1324 4 US-09-645-791-13
35 177.5 4.5 3248 1 US-08-353-700-1
36 177.5 4.5 3248 5 PCT-US95-16216-1
37 177 4.5 1805 1 US-07-853-913-2
38 177 4.5 8991 4 US-08-714-741-32
39 176.5 4.5 1400 1 US-08-080-255-7
40 176.5 4.5 1400 3 US-08-455-713-7
41 176.5 4.5 1400 5 PCT-US93-05857-7
42 174.5 4.4 1588 5 PCT-US93-07261-11
43 174.5 4.4 1663 5 PCT-US93-07261-16
44 174 4.4 1354 3 US-08-685-871-2
45 173.5 4.4 1298 4 US-09-645-456A-14


```
NAME/KEY: CARBOHYD
LOCATION: 1226..1228
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 102..105
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 663..666
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 808..811
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 885..888
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 17..19
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 31..33
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 41..43
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 100..102
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 140..142
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..218
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 471..473
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 507..509
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 531..533
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 591..593
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 656..658
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 801..803
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 812..814
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 815..817
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 876..878
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 888..890
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 939..941
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1060..1062
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1128..1130
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1129..1131
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1135..1137
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1181..1183
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1208..1210
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1249..1251
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 47..50
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 126..129
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 157..160
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 158..161
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 159..162
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..219
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 274..277
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 276..279
```



```

; TITLE OF INVENTION: BIOSYNTHESIS PATHWAY
; FILE REFERENCE: 4859-0027-0 US/09/282,146A
; CURRENT APPLICATION NUMBER: US/09/282,146A
; CURRENT FILING DATE: 1999-03-31
; EARLIER APPLICATION NUMBER: JP 10-125171
; EARLIER FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-09-282-146-2

Query Match      5.4%; Score 210.5; DB 4; Length 200;
Best Local Similarity 30.7%; Pred. No. 1.4e-09;
Matches 39; Conservative 26; Mismatches 41; Indels 21; Gaps 2;

QY 384 APARETCVECKTVYPMERLLANQQVFIHSCFRCSYCNKLSLGTYSALHGRYCKPHEN 443
Db 4 AGTTQKMACDKTVYLVLDKLTADNRVFKACFRCHHCKGTGVLKLNYSNPFVLYCRPHFD 63
QY 444 QLFKSKGNVDEGF-----GHRPHKDLWASKNENEIILER-----PAQLA 482
Db 64 QLFKQTSLDKSPGTPKXVVKPKPIDSEKPVAKVTSMPGGTREKCFGCKKTVYITEKV 123
QY 483 NARETTPH 489
Db 124 SANGTPY 130

RESULT 4
US-08-934-627B-2
; Sequence 2, Application US/08934627B
; Patent No. 6159174
; GENERAL INFORMATION:
; APPLICANT: OSAMU HASIGAWA
; APPLICANT: SATOSHI AOTSUKA
; APPLICANT: SOICHIRO TAKENISHI
; APPLICANT: HIROFUMI UCHIMIYA
; TITLE OF INVENTION: COTTON PLANT GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 3424 Peachtree Road, N.E., 2400 Monarch Tower
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,627B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Roger T. Frost
; REGISTRATION NUMBER: 22,176
; REFERENCE/DOCKET NUMBER: 20111-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 949-2400
; TELEFAX: (404) 949-2499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
US-08-934-627B-2

Query Match      5.2%; Score 205.5; DB 3; Length 208;
Best Local Similarity 35.2%; Pred. No. 3.9e-09;
Matches 44; Conservative 21; Mismatches 59; Indels 1; Gaps 1;

QY 379 MKKFOAPARETCVECKTVYPMERLLANQQVFIHSCFRCSYCNKLSLGTYSALHGRYIC 438
Db 1 MATFQG-TQKCNACNKTIVYLVLDKLTADNRVFKACFRCHHCKGTGTLKLSNYSNPFVLYC 59
QY 439 KPHFNQLEKSKGNVDEGFHGRPHKDLWASKNENEIILERPAQLANARETTPHSPGVEDAPI 498
Db 60 RHYDQLFRTGSLDKSPGTPKVVKPERQIDSSALKVMNSFGGTREKCAACKTAYPI 119
QY 499 AKGVV 503
Db 120 ERVTV 124

RESULT 5
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/741,940
; APPLICATION NUMBER: 19920109
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-07-741-940-7

```

```
Query Match      4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKNKSAIVEIFSKYKAAEET---NMEKKRSNTENLSQHFR----- 65
DB 1788 VRKNAD-----SKNLNARVFSNDKSKQNLKNNKDFNDKLPNNEDRVGSAFD 1840
QY 66 -----KGTTLVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPAE--VTSH---A 112
DB 1841 SPHHYPIEGTPTVCFRNSDLSLDDDDDDVLSREKAEELRKAKENKESEAKVTSHTELT 1900
QY 113 ASGAKADQEQIHPRSRLRSPPEALVQGRYPHIKQGEDLKDQ--STESKKMEN----- 163
DB 1901 SNOQSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPIRGAATDEKLQFAIENTPV 1960
QY 164 CLGESRHEVEKSEISNTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
DB 1961 CFSHNSLSLSLSDIQENN-----NKNEPIKETEPPDSQGEPSKQ-----ASG-- 2005
QY 224 ISENSYSLLDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSK 280
DB 2006 YAPKSFHVEDTPVCFRNSLSLSLSDSEDD-----LQECISSAMPK-----KK 2050
QY 281 QSSSTNTYNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
DB 2051 KPSRLKGDNEKISPRNMGILGEDLTLDLKDQRPDSEHGLSPDSE---NFDWKAIOEG 2106
QY 329 ENSL-----AVRSTPAEDSDRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
DB 2107 ANSIVSSLHCAAAAACLSRQASDS--DSILSLKSGISLSPFH---LTPDQEKPTSNK 2162
QY 375 PPKAMKKFQAPARETCVQCKTVYPMERLLANQQVPHISCFCRSCYNNKLSLGTVASLHG 434
DB 2163 GPRILK---PGEKSTLETK----- 2179
QY 435 RIYCKPHFNOLFKSKNYDEGGHRPHKDLWASK-NENBEI---LERPAQ-----LANAR 485
DB 2180 -----IESESKGIRG-GKKVYKSLITCKVRNSEISGQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVLAASMEAKASSQOE-----KEDKPAETKKLRIANPP 534
DB 2228 TWIHIPGVNSSSSTSPVSKGPPLKTPASKSPSEGOATTTPRGAKEVSKSELSFVARQ 2287
QY 535 PTELGGSSGA-LEEGTKMKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERS--RPF 591
DB 2288 TSQIGSSKAPSRSGSRDSTPSRPAQOPLSRP-----IQSPGRNSISPRNGISPP 2338
QY 592 TAAASFQSTSVKSPKTVSPPIKAGWSMSEQSESVGGRVAERKQVENAKSKNGNNGVKT 651
DB 2339 NXLSQLPRTS--SPSTAS---TKSSGSGKVSYSFQRMQSQOQLTKQTLGSKNASSIPRS 2393
QY 652 TWQNKESKGETGKRSKEGHSLEMENL---VENGADSDSDNSFLKQOS 698
DB 2394 ---ESASKGLNQMNNGANKKVELSRMSTKSSGSESDRSPRLVLRQS 2440

RESULT 6
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
```

```
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-289-548A-7

Query Match      4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKNKSAIVEIFSKYKAAEET---NMEKKRSNTENLSQHFR----- 65
DB 1788 VRKNAD-----SKNLNARVFSNDKSKQNLKNNKDFNDKLPNNEDRVGSAFD 1840
QY 66 -----KGTTLVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPAE--VTSH---A 112
DB 1841 SPHHYPIEGTPTVCFRNSDLSLDDDDDDVLSREKAEELRKAKENKESEAKVTSHTELT 1900
QY 113 ASGAKADQEQIHPRSRLRSPPEALVQGRYPHIKQGEDLKDQ--STESKKMEN----- 163
DB 1901 SNOQSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPIRGAATDEKLQFAIENTPV 1960
QY 164 CLGESRHEVEKSEISNTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
DB 1961 CFSHNSLSLSLSDIQENN-----NKNEPIKETEPPDSQGEPSKQ-----ASG-- 2005
QY 224 ISENSYSLLDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSK 280
DB 2006 YAPKSFHVEDTPVCFRNSLSLSLSDSEDD-----LQECISSAMPK-----KK 2050
QY 281 QSSSTNTYNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
DB 2051 KPSRLKGDNEKISPRNMGILGEDLTLDLKDQRPDSEHGLSPDSE---NFDWKAIOEG 2106
QY 329 ENSL-----AVRSTPAEDSDRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
DB 2107 ANSIVSSLHCAAAAACLSRQASDS--DSILSLKSGISLSPFH---LTPDQEKPTSNK 2162
QY 375 PPKAMKKFQAPARETCVQCKTVYPMERLLANQQVPHISCFCRSCYNNKLSLGTVASLHG 434
DB 2163 GPRILK---PGEKSTLETK----- 2179
QY 435 RIYCKPHFNOLFKSKNYDEGGHRPHKDLWASK-NENBEI---LERPAQ-----LANAR 485
DB 2180 -----IESESKGIRG-GKKVYKSLITCKVRNSEISGQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVLAASMEAKASSQOE-----KEDKPAETKKLRIANPP 534
DB 2228 TWIHIPGVNSSSSTSPVSKGPPLKTPASKSPSEGOATTTPRGAKEVSKSELSFVARQ 2287
QY 535 PTELGGSSGA-LEEGTKMKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERS--RPF 591
DB 2288 TSQIGSSKAPSRSGSRDSTPSRPAQOPLSRP-----IQSPGRNSISPRNGISPP 2338
QY 592 TAAASFQSTSVKSPKTVSPPIKAGWSMSEQSESVGGRVAERKQVENAKSKNGNNGVKT 651
DB 2339 NXLSQLPRTS--SPSTAS---TKSSGSGKVSYSFQRMQSQOQLTKQTLGSKNASSIPRS 2393
QY 652 TWQNKESKGETGKRSKEGHSLEMENL---VENGADSDSDNSFLKQOS 698
DB 2394 ---ESASKGLNQMNNGANKKVELSRMSTKSSGSESDRSPRLVLRQS 2440
```

Db 2163 GPRILK-----PGEKSTLETKK----- 2179
QY 435 RIYCKPHNFQKSGNYDEGHRPHKDLWASK-NENEI-----LERPAQ-----LANAR 485
Db 2180 -----IESEKGIKG-GKKVYKSLITGKVRNSSEISQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAWPP 534
Db 2228 TWIHIPGVNRSSSTSPVSKGPPKLTTPASKSPSEGQTATTSRGAKEPVSKSELSPVARQ 2287
QY 535 PTELGGSSGA-LEEGIKMVKPKWPPDEISKPEVPEDVDLKKLRSSSLKERS--RPF 591
Db 2288 TSQIGSSKAPSRGSRDSTPSRPAQPLSRP-----ISPGNSISPGNGISPP 2338
QY 592 TVAASFQSTSVKSPKTVSPPIRKGWMSSEQSESVGGRVAERKQVENAKAKKXGNGVCKT 651
Db 2339 NKLSOLPRTS--SPSTAS---TKSSGSKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL---VENGADSDDDNSFLKQOS 698
Db 2394 ---ESASKGLNQMNNGANGANKKVELSRMSSTKSGSSEDRSERPVLVRQS 2440

RESULT 7
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVARIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7
Query Match 4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;
QY 26 VNKXSSAIVFISFYKQKAAEET---NMEKRSNTENLSQHFR----- 65
Db 1788 VRKNAD-----SKNLAERVFSDNKDKSKQNLKNNKDFNDKLPNNEDRVGRSFAFD 1840
QY 66 -----KGLTVLKKWENPGJCAESHTDSLANSSTEIRHADHPAE--VTSH---A 112
Db 1841 SPHHYTPTEGTPCYCFSRNDSLSLDFDDDDVDLSREKAEELKAKENKSEAKVTSHTLT 1900
QY 113 ASGAXADEQEIHPHRSRLRSPPEALVQGRYPHIKDGEDLKH-STEKXKVEN----- 163
Db 1901 SNQGSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPIRGAATDEKLQNFATENTPV 1960
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKEGPEPTQTKILRAQSASGRK 223
Db 1961 CFSHNSLSLSLSDIQENN-----NKNEPIKETETPPDSQGEPSKPQ-----ASG-- 2005
QY 224 ISENSYSLDDLEI---GPGQLSSSTFDSEKNSRRNLELPRLSETSIKDRMAKYQAAVSK 280
Db 2006 YAPKSFHVEDTPCYFSRNSLSUSSISDSEDD-----LLOECISSAMPK-----KK 2050
QY 281 QSSSTNYTNEL---KASGG-----BIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
Db 2051 KPSRLKGDNRKXSPNMGGLIGEDLTLDLKDIOQPDSEHGLSPDSE-----NFDWKAQEG 2106
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDGRASSLSSES 374
Db 2107 ANSIVSSLHQAAAAACLRQASDS--DSILSLKSGISLGSFFH---LTPDOEKPFSTGNK 2162
QY 375 PPKAMKKECAPARETCVCOQTVYPMERLLANQOVFHISCFRCSCYCNKLSLGTIVASLHG 434
Db 2163 GPRILK---PGEKSTLETKK----- 2179
QY 435 RIYCKPHNFQKSGNYDEGHRPHKDLWASK-NENEI---LERPAQ-----LANAR 485
Db 2180 -----IESEKGIKG-GKKVYKSLITGKVRNSSEISQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAWPP 534
Db 2228 TWIHIPGVNRSSSTSPVSKGPPKLTTPASKSPSEGQTATTSRGAKEPVSKSELSPVARQ 2287
QY 535 PTELGGSSGA-LEEGIKMVKPKWPPDEISKPEVPEDVDLKKLRSSSLKERS--RPF 591
Db 2288 TSQIGSSKAPSRGSRDSTPSRPAQPLSRP-----ISPGNSISPGNGISPP 2338
QY 592 TVAASFQSTSVKSPKTVSPPIRKGWMSSEQSESVGGRVAERKQVENAKAKKXGNGVCKT 651
Db 2339 NKLSOLPRTS--SPSTAS---TKSSGSKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL---VENGADSDDDNSFLKQOS 698
Db 2394 ---ESASKGLNQMNNGANGANKKVELSRMSSTKSGSSEDRSERPVLVRQS 2440

RESULT 8
US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA

HEDGE, PHILIP J.
JOSLYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAKAMURA, YUSUKE
THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CLONE: APC
IMMEDIATE SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7
Query Match 4.9%; Score 193; DB 4; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;
26 VNKNKSSAIVEIFSYQKAAET---NNEKKRSNTENLSQHFR----- 65
1788 VRKNAD-----SKRNLAERVFSDNKKDKQNKQNKNDKLPNNEDVRGSPAFD 1840
66 -----KGLTVLKKXWENPGLGAESHTDSLRNSSTEIRHRADHPAE--VTSH---A 112
1841 SPHHVTPTEGTYCYFSRNDLSLDFDDDDVDLSREKAEKAKENKESEAKVTSHTLT 1900
113 ASGAKADQEQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHP-STESKKMEN----- 163
1901 SNQQSANKTQALAKQPINRGQPKPILOQSTFPQSSKDIPDGAATDEKLQNFATENTPV 1960
164 CLGESRHEVEKEISENTDASKIEKYVPLNRLKMFKEGPEPTQTKILRAQSRSAGRK 223
1961 CFSHNSLSLSLSDIDQENN-----NKENEPKETEPPDSQGEPSKPQ-----ASG-- 2005
224 ISENSYSLDDLEI---GFCQLSSSTFDSEKESRRNLPLRLSEISIKDRMAKYQAIVSK 280
2006 YAPKGFHVEDTPVCFSRNSSSSLSIDSEDD-----LLOECISSAMPK-----KK 2050

281 QSSSTNYTNEL---KASGG-----EIK-IHKMEKENVPPGPEVCITHQGEKISAN 328
2051 KPSRLKGDNEKHSRPNMGILGEDLTLDLKDIOQPDSEHGLSPDSE-----NFDWKAIOEG 2106
329 ENSL-----AVRSTPAEDDSRDS--QVKEVQ--QPVHPKPLSPDSRASSJESS 374
2107 ANSIVSSLHQAAAAACLSROASSDS-DSILSLKSGISLGSPPH---LTPQEEKPFTSNK 2162
375 PPRAMKXFOAPARETCVCEOKTVYPMERLLANOQVPHISCFRCSCYCNKLSLGTIASLHG 434
2163 GPRILK-----PGEKSTLETKK----- 2179
435 RIYCKPHFNOLFYSKGNVDEGFGHRPHKDIWASK-NENEIE---LBRPAO-----LANAR 485
2180 -----ISESKGIKG-GKKVYKSLITKVRNSNISEIGMKOPLQANMPSISRGR 2227
486 ETPHSPGVED-----APIAKVGVLAASMEAKASQOE-----KEDKPAETTKLRIAWPP 534
2228 TMIHPCVRNRSSTSPVSKKGPLKTPASKSSEGTATTSRGAKPSVKSELSPVARQ 2287
535 PTELGSSGSA-LEEGIKMKPKWPPDEISKPEVPDVLDLKKLRSSSLKERS--RPF 591
2288 TSQIGGSSKAPSRSGSRDSTPSRPAQOPLSRP-----IQSPGRNSISPGRNGISPP 2338
592 TVAASFQSTSVKSPKTVSPPIRKGMSESESESVGGRVAERKQVENAKASKKNGNVGKT 651
2339 NKLSQLPRTS--SPSTAS---TKSSGSGKMSYSPGRQMSQQNLTKQTGLSKNASSIPRS 2393
652 TWQNKESKGTGKSKEGHGLEMENENL---VENGADSDDEDDNSFLKQOS 698
2394 ---ESASKGLNQMNNGANGANKVELSRMSSTKSGSESRSERPVLVRQS 2440
RESULT 9
US-07-741-940-2
Sequence 2, Application US/07741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-741-940-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
 Best Local Similarity 19.4%; Pred. No. 2e-06;
 Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKKNSSAIVFISKYQAAEET---NMKKRSNTENLSQHFR----- 65
 DB 1789 VRKNAD-----SKNLLNARVFSNDKSKQNLKNNKDFNDKLPNNEDRVGSGAFD 1841
 QY 66 -----KGLTLVKKWNPGLGAESHTDLSRNSSTEIRHRADHPAE--VTSH---A 112
 DB 1842 SPHYTPTEGTPCFSRNDSLSLDPDDDDVLSREKAEELKAKENKESAKVTSHTLT 1901
 QY 113 ASGAKADQEQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDQ--STESKKMEN----- 163
 DB 1902 SNOOSANKTOAIKQPINRGQPKPILOKQSTFPQSKDIPDRGAATDEKLNFAIENTPV 1961
 QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRASGRK 223
 DB 1962 CFHNSLSLSLSDIDQENN-----NKNEPIKTEPPDQSGEPEKQP-----ASG-- 2006
 QY 224 ISENSYSLDDLEI---GPGQLSSSTFSEKNSRRNLELPRLSETIKDRMAKYQAAVSK 280
 DB 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LLOECISSAMPK-----KK 2051
 QY 281 QSGSTNYTNEL---KASGG-----EIK-IHKMEQKENVPGPRVCITHQEGEKISAN 328
 DB 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE---NFDWKAIQEG 2107
 QY 329 ENSL-----AVRSTPAEDSDRS--QVSEVQ--QPVHPKPLSPDSRASSLSSESS 374
 DB 2108 ANSTVSSLHQAAAAACLSRQSSDS--DSILSLKSGISLGSPFH---LTPDQEEKPFTSNK 2163
 QY 375 PPKAMKXFOAPARETCVECKTVPMERLLANQOVPHISCFRCSCYCNKLSLQTYASLHG 434
 DB 2164 GPRILK---PGEKSTLETK----- 2180
 QY 435 RIYCKPHFNQLFKSGNYDEGFGRRPHKDLWASK--NENEEI---LERPAQ-----LANAR 485
 DB 2181 -----IESESKGIGK-GKKVYKSLITGKVRNSNLSIQMKQPLQANWPSISRGR 2228
 QY 486 ETHSPQVED-----APIAKGVULAAAMEAKASQOE-----KEDKPAETKKLRIAWPP 534
 DB 2229 TMIHIPGVRRSSSTSPVSKKGPPLKTPASKSPSEGGTATTSRPAKSPVSKSELSPVARQ 2288
 QY 535 PTELGGSSGA--LEEGIKMSKPKWPPDEISKPEVPEDVDLDLKLRRSSSLKERS--RPF 591
 DB 2289 TSQIGGSKAPSRSGSRDSTSPRAQOPLSRP-----IOSPGNSISPGNRGSRPP 2339
 QY 592 TVAASFQSTSVKSPKTPVPIRKGWMSSEQSEBSVGGRVAERQVENAKKNGNVGKT 651
 DB 2340 NKLSQLPRTS--SPSTAS---TKSSGSKMSYTPSGRQMSQQLNTQTGLSKNASSIPRS 2394
 QY 652 TWNKESKGTGRKSGEGHSELENNENL-----VENGADSDDEDSFLKQOS 698
 DB 2395 ---ESASKGLNQMNGNGANKKVELRMSSTKSGSSEDSRSPVLRQS 2441

RESULT 10
 US-08-289-548A-2
 Sequence 2, Application US/08289548A
 Patent No. 5646212
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, LTD
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/289,548A
 FILING DATE: 12-AUG-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107,46943
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-289-548A-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
 Best Local Similarity 19.4%; Pred. No. 2e-06;
 Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKKNSSAIVFISKYQAAEET---NMKKRSNTENLSQHFR----- 65
 DB 1789 VRKNAD-----SKNLLNARVFSNDKSKQNLKNNKDFNDKLPNNEDRVGSGAFD 1841
 QY 66 -----KGLTLVKKWNPGLGAESHTDLSRNSSTEIRHRADHPAE--VTSH---A 112
 DB 1842 SPHYTPTEGTPCFSRNDSLSLDPDDDDVLSREKAEELKAKENKESAKVTSHTLT 1901
 QY 113 ASGAKADQEQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDQ--STESKKMEN----- 163
 DB 1902 SNOOSANKTOAIKQPINRGQPKPILOKQSTFPQSKDIPDRGAATDEKLNFAIENTPV 1961
 QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRASGRK 223
 DB 1962 CFHNSLSLSLSDIDQENN-----NKNEPIKTEPPDQSGEPEKQP-----ASG-- 2006
 QY 224 ISENSYSLDDLEI---GPGQLSSSTFSEKNSRRNLELPRLSETIKDRMAKYQAAVSK 280
 DB 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LLOECISSAMPK-----KK 2051
 QY 281 QSGSTNYTNEL---KASGG-----EIK-IHKMEQKENVPGPRVCITHQEGEKISAN 328
 DB 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE---NFDWKAIQEG 2107
 QY 329 ENSL-----AVRSTPAEDSDRS--QVSEVQ--QPVHPKPLSPDSRASSLSSESS 374
 DB 2108 ANSTVSSLHQAAAAACLSRQSSDS--DSILSLKSGISLGSPFH---LTPDQEEKPFTSNK 2163

375 PPKAMKFFQAPARETCVECKTYPYMERLLANQOVPHISCFRCSYCNKLSLGTASYLHG 434
2164 GPRILK-----PGEKSTLETKK----- 2180
435 RIYCKPHFNQLFKSGNYDEGFGHRRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
2181 -----IESEKGIKG-GKKVYKSLITGKVRNSNISESQMKQPLQANWPSISRGR 2228
486 ETPHSGVED-----APIAKVGVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
2229 TMIHIGVNRNSSSTSPVSKGKPLTPASKSPSEGOATTTPRGAKPSVKSELSPVARQ 2288
535 PTELGGSSGA-LEEGIKMSPKPPPEDEISKPEVPEDVDLDLKKURRSSLSLKERS--RPF 591
2289 TSQIGSSKAPSRSGSRDSTPSRPAQPLSRP-----IQSPGRNSISPGRNGISPP 2339
592 TVAAFSQTSVKSPKTVSPPIRKGMSEOSSESVGGRVAERKQVENAKSKKNGNVGKT 651
2340 NKLSQLPRTS--SPSTAS---TKSSSGKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2394
652 TWQNKESKGETGKRSKEGHSLEMEENL---VENGADSDDDNSFLKQOS 698
2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSSGESDRSERPVLVRQS 2441

RESULT 11

US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids

TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-452-654-2
Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 293; Indels 214; Gaps 37;
QY 26 VNKKSSAIVEIFSKYKAAEET---NMEKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNNLNAERVFSDNKKSKQNLKNNKDFDKLPNNEDRVGRGSPAFD 1841
QY 56 -----KGTLTVLVKKWENPGLGAESHTDSLNSSTEIRHADHPAE--VTSH---A 112
Db 1842 SPHYPTIEGTYPFCFSRNDLSLQFDDDDVDLSREKAEELKAKENKESAEKVTSHTELT 1901
QY 113 ASGAKADEEIGHPSRLRSPREALVOGRYPHIKDGDLKDH-STESKCMEN----- 163
Db 1902 SNOQSAKQTAIAKQPINRGQPKFILOKQSTFPOSSKDIPDRGAATDEKLNFAIENTPV 1961
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLUMKMFKEGTEPTQTKILRAQSASGOK 223
Db 1962 CFSHNSLSLSLSDIDQENN-----NKENEPKETETPPDSQGEPSKPO-----ASG-- 2006
QY 224 ISENSYSLDDLEI---GPGQLSSSTFDSERKNESRNLELPRLSETSIKDRMAKYQAAVSK 280
Db 2007 YAPKSFHVEDTPVCFSRNSSLSSLSIDSEDD-----LLQECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEI---KASGG-----EIK-IHKMEOKENVPPGPEVCITHQEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEDTLDLKDIOQPDSEHGLSPDSE---NFDWKATQEG 2107
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAACLSRQASSDS-DSILSLKSGISLGSFFH---LTPDOEEKPFTSNK 2163
QY 375 PPKAMKFFQAPARETCVECKTYPYMERLLANQOVPHISCFRCSYCNKLSLGTASYLHG 434
Db 2164 GPRILK-----PGEKSTLETKK----- 2180
QY 435 RIYCKPHFNQLFKSGNYDEGFGHRRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----IESEKGIKG-GKKVYKSLITGKVRNSNISESQMKQPLQANWPSISRGR 2228
QY 486 ETPHSGVED-----APIAKVGVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIGVNRNSSSTSPVSKGKPLTPASKSPSEGOATTTPRGAKPSVKSELSPVARQ 2288
QY 535 PTELGGSSGA-LEEGIKMSPKPPPEDEISKPEVPEDVDLDLKKLRSSLSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGSRDSTPSRPAQPLSRP-----IQSPGRNSISPGRNGISPP 2339
QY 592 TVAAFSQTSVKSPKTVSPPIRKGMSEOSSESVGGRVAERKQVENAKSKKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSSGKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2394
QY 652 TWQNKESKGETGKRSKEGHSLEMEENL---VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSSGESDRSERPVLVRQS 2441

RESULT 12

US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH
APPLICANT: MAKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-655B-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNNKSSAIVEISKYKAAET---NMEKRSNTENLSQHFR----- 65
DB 1789 VRKNAD-----SKNLAERVFSDNKKQKQNLKNSKDFNDKLPNNEDRVGSAFD 1841
QY 66 -----KGLTVLKKWENPGLGAESHTDLSRNSSTEIRHRADHPAE--VTSH---A 112
DB 1842 SPHHYPIEGTPTCFERNDSLSLDFDDDDVLSREKAEIRAKENKESAEKVTSHTELT 1901
QY 113 ASGAKADQEEQIHPRSLRSPPEALVQGRYPHIKDGEDLKDH-STESKKWEN----- 163
DB 1902 SNOQSANKTQAIKQPIKQKSTPQSSKIDPDRGAATDEKLQFAIENTPV 1961
QY 164 CLGESHEVEKSEISNTDASGIEKYVPLNKLKMWFKGSEFTQTKILRAQSRASGRK 223
DB 1962 CFSHNSLSLSLSDIDQENN-----NKENEPKETEPPDSQGEPSKPQ-----ASG-- 2006
QY 224 ISENSYSLLDLEI---GPOQLSSTFDSEKNESRRNLELPRISEISIKDRMAKYQAAYSK 280
DB 2007 YAPKSFHVEDTPTCFERNDSLSLSLSDSEDD-----LQECISAMPK-----KK 2051
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEKENVPPGPEVCITHOGEKISAN 328
DB 2052 KPSRLKGDNEKHSRNMWGLGEDLTLDLKDQRPDSEHGLSPDSE-----NFDWKAQEG 2107
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QVHPKPLSPDSRASSLSSESS 374

DB 2108 ANSIVSSLHQAAAAACLSRQASDS-DSILSKSGISLGSFPH---LTPDOEEKPFTSNK 2163
QY 375 PKAMWKQFAPARETCVQCQTVPMERLLANQOVFHIHSCFRCYCNNKLSLGTYSALHG 434
DB 2164 GPRILK-----FOEKSTLETKK----- 2180
QY 435 RIYCKPHFNQLFKSGNYDEGFGHRPHKDLWASK-NEVEEI---LERPAQ-----LANAR 485
DB 2181 -----IESEKGIKG-GKKVYKSLITGVRNSNISEISGQMKQPLQANMPSISRGR 2228
QY 486 ETPHSPGVED-----APIAKVGLAASMEAKASQOE-----KEDKPAETKKLRIAMPP 534
DB 2229 TMIHIPGVNRSSSTSPVSKKGPPLKTPASKSPSEGQTATTPRGAKPSVKSELSPVARQ 2288
QY 535 PTELSSGSA-LEEGIKMSKPKWPPDEISKPEVPEDVDLDLKLRRSSSLKERS--RPF 591
DB 2289 TSQIGSSKAPSRGSRDSTSRPAQQLSRP-----IQSPGRNSISPGRNGISPP 2339
QY 592 TVAASFQTSVKSPKTVSPPIRKGWSMSEOSEESVGGRVARERKOVENAKAKKNGNVGKT 651
DB 2340 NKLSQLPRTS--SPSTAS---TKSSGSGKMSYTPGRQMSQQLTKQTGLSKNASSIPRS 2394
QY 652 TWQNKESKGETGKRSKEGHSLEMEENL---VENGADSDDEDNSFLKQOS 698
DB 2395 ---ESASKGLNQMNGNGANKKVELSRMSSTKSSGSESDRSERPVLVRQS 2441

RESULT 13
US-08-452-655B-7
Sequence 7, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MAKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-452-655B-7

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKXSSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
Db 1789 VRKVAD-----SKNUNLAERVFSKQKQNLKNNKSKDFNDKLPNNEDVRGSAFDP 1841

QY 66 -----KGTITLVKKWENPGLGAESHTDLSRNSSTEIRHRADHPAE--VTSH---A 112
Db 1842 SPHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAELEKAKENKESAKVTSHTELT 1901

QY 113 ASKADQEQIHPRLSRRLSPPEALVQGRYPHDKGEDLKH-STESKKNEN----- 163
Db 1902 SNOQANKTOAIKQPIRNGQPKPIQKSTFPQSSKDIPDGAATDEKLFQFAIENTPV 1961

QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLSDIQENN-----NKNEPIKETETPPDSQGEPSKPQ-----ASG-- 2006

QY 224 ISENSYSLDDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETSIKDRMAKYQAAVSK 280
Db 3007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LQECISSAMPK-----KK 2051

QY 281 QSSSTNYTNEL---KASG-----EIK-IHKMEQKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107

QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2180 -----PCEKSTLETCK----- 2180

QY 435 RIYKPHFNQLEKSKGNVDGEGHPRPHKDLWASK-NENEEL---LERPAQ-----LANAR 485
Db 2181 -----IESESKGIGK-GKVKYKSLITGKVRNSSEISGQMKPLQANMPFSISGR 2228

QY 486 ETPHSPGVED---APIAKGVLAASMEAKASSQOE-----KEDKPAETKRLIAWPP 534
Db 2229 TWIHIPGVNSSSTSPVSKGPPKTPASKPSSEGQTATSPRGAKPSVKSLSPLVARQ 2288

QY 535 PTELSSGSA-LEEGIKMSKPKWPPEDISKEPEVEDVDLKLRRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRGSRDSTSPRPAQPLSRP-----IQSPGRNISIPIGANGISPP 2339

QY 592 TVAASFQSTSVKSPKTPVSPPIRKGNMSGEQEEESVGGRAVERKQVENAKSKXGNVGKT 651
Db 2340 NKLSQLPRTS--SPETAS---TKSSGSGKMSYTPSGRQMSQOQLTKQTGLSKNASSIPRS 2394

QY 652 TWQNKESKETGRSKEGHSLEMEENEL---VENGCADSDDEDNLFKQOS 698
Db 2395 ---ESAKGLQNMNGNGANKKVELLSRMSSTKSGSSESDRSERPVLVRQS 2441

RESULT 14

US-08-370-235A-2

; Sequence 2, Application US/08370235A

Patent No. 5910418
GENERAL INFORMATION:
APPLICANT: VOGELSTEIN, BERT
APPLICANT: KINZLER, KENNETH W.
APPLICANT: HILL, DAVID E.
APPLICANT: JOHNSON, KAREN A.
TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & WITCOFF, LTD.
STREET: 1001 G STREET, N.W.
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,235A
FILING DATE: 01-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.48688
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 508 9100
TELEFAX: 202 508 9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-370-235A-2

Query Match 4.9%; Score 193; DB 2; Length 2843;

Best Local Similarity 19.4%; Pred. No. 2e-06;

Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKXSSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
Db 1789 VRKVAD-----SKNUNLAERVFSKQKQNLKNNKSKDFNDKLPNNEDVRGSAFDP 1841

QY 66 -----KGTITLVKKWENPGLGAESHTDLSRNSSTEIRHRADHPAE--VTSH---A 112
Db 1842 SPHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAELEKAKENKESAKVTSHTELT 1901

QY 113 ASKADQEQIHPRLSRRLSPPEALVQGRYPHDKGEDLKH-STESKKNEN----- 163
Db 1902 SNOQANKTOAIKQPIRNGQPKPIQKSTFPQSSKDIPDGAATDEKLFQFAIENTPV 1961

QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLSDIQENN-----NKNEPIKETETPPDSQGEPSKPQ-----ASG-- 2006

QY 224 ISENSYSLDDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETSIKORMAKYQAAVSK 280
Db 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LQECISSAMPK-----KK 2051

QY 281 QSSSTNYTNEL---KASG-----EIK-IHKMEQKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107

QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQASDS--DSLILSKSGISLSPFH---LTPDQEKPTSNK 2163

QY 375 PPXAMKFKQAPARETCVEQKTVYPMERLLANQQVPHISCFRCSYCNKNKLSLGTASLHG 434

Db 2164 GPRILK-----PGSKSTLETK----- 2180
Qy 435 RIYCKPHNFQKSKGNVDEFGHPRHKLWASK-NENEEI-----LERPAQ-----LANAR 485
Db 2181 -----IESESKGIGK-CKVKYKSLITCKVRSNSENSEISQMKQPLQANMPSISGR 2228
Qy 486 ETHSPQVED-----APIAKGVVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIPGVNRSSSTSPVSKGPKPLKTPASKSPSEGGTATTSPRGAKPSVKSELSPVARQ 2288
Qy 535 PTELGGSSGA-LEEGIKMSPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGRSDTSPSPAQQPLSRP-----IOSPGNLSISPGNIGISPP 2339
Qy 592 TVAASFQSTSVKSPKTVPPPIRKGMNSQEESVGGVRAERKQVENAKASKKNGVNGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSGKMSYTPSGRQMSQQNLTKQTGLSKNASSIPRS 2394
Qy 652 TWONKESKGTGKRSKEGHSLEMNENL-----VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSGSSESDRSERPVLVRQS 2441

RESULT 15

US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MAKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-582-2
Query Match 4.9%; Score 193; DB 3; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;
Qy 26 VNKKSSAIVIFSKYQAAEET---NMEKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNNLNAERFSDNKKSKQNLKNSKDFNDKLPNNEDVRGSAFD 1841
Qy 66 -----KGTLLVLKKWENPGIAGSHTDSLNSSTEIRHRADHPAE--VTSH---A 112
Db 1842 SPHYPTPIEGTPYCFSRNDSLSLDFDODDDVDLSREKAELKAKENKESAKVTSHTLT 1901
Qy 113 ASGAKADQEEQIHPRSLRSPPEALVQGRYPHIKDGEDLDKH-STESKKMEN----- 163
Db 1902 SNOQSANKTQAIKQPINRGQPKPILOKQSTFPOSSKDIIDRGAATDEKLQFAIENTPV 1961
Qy 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSASGRK 223
Db 1962 CFSHNSLSLSLSDIDQENN---NKENERIKETEPPDSQSQEPKQP-----ASG-- 2006
Qy 224 ISENSYSLDDLEI---GPGQLSSSTFSEKNSRNLPLSLSETSIKDRMAKYQAASVK 280
Db 2007 YAPKSFHVEDTFVCFSRNLSLSLSDSEDD-----LLOECISSAMPK-----KK 2051
Qy 281 OSSSTNYTNEL---KASGG-----EIK-IHKVEOKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEBLTLDLKIQPDSEHGLSPDSE-----NFDKAIQEG 2107
Qy 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQASDS--DSILSKSGISLGSPPH---LTPDQEEKPFTSNK 2163
Qy 375 PPKAMKFOAPARETCVECOQTVPMERLLANQQVFHISCPCRCYCNNKLSLGTAYASLHG 434
Db 2164 GPRILK-----PGKSTLETK----- 2180
Qy 435 RIYCKPHNFQKSKGNVDEFGHPRHKLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----IESESKGIGK-GKKVYKSLITGKVRNSENSEISQMKQPLQANMPSISGR 2228
Qy 486 ETHSPQVED-----APIAKGVVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIPGVNRSSSTSPVSKGPKPLKTPASKSPSEGGTATTSPRGAKPSVKSELSPVARQ 2288
Qy 535 PTELGGSSGA-LEEGIKMSPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGRSDTSPSPAQQPLSRP-----IOSPGNLSISPGNIGISPP 2339
Qy 592 TVAASFQSTSVKSPKTVPPPIRKGMNSQEESVGGVRAERKQVENAKASKKNGVNGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSGKMSYTPSGRQMSQQNLTKQTGLSKNASSIPRS 2394
Qy 652 TWONKESKGTGKRSKEGHSLEMNENL-----VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSGSSESDRSERPVLVRQS 2441

Search completed: January 6, 2004, 09:48:58
Job time : 25 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:47:24 ; Search time 38 Seconds
(without alignments)
4007.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRRQWTSLSLRVTA.....LSVEQIKENRYDEDEEE 759

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*

3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*

7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*

10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*

12: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*

13: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*

14: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*

15: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*

16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3869	98.5	755	11	US-09-783-732-4
2	3027.5	77.1	596	11	US-09-783-732-2
3	1254	31.9	243	9	US-09-925-297-611
4	464.5	11.8	527	9	US-09-880-192-53
5	464.5	11.8	527	12	US-10-427-348-53
6	363	9.2	158	9	US-09-789-919-42
7	357	9.1	127	12	US-10-440-366-12
8	357	9.1	127	12	US-10-144-047-2959
9	281	7.2	52	11	US-09-783-732-5
10	245	6.2	1965	12	US-10-359-012-6
11	223	5.7	1312	12	US-10-136-704-29
12	223	5.7	1312	15	US-10-071-179-29
13	222	5.7	2476	11	US-09-824-574-7
14	221	5.6	4723	12	US-10-359-012-8
15	213	5.4	1633	12	US-10-359-012-4

Query Match	Score	Length	ID	Description
Sequence 4, Appl	98.5%	755	11	US-09-783-732-4
Sequence 2, Appl	77.1%	596	11	US-09-783-732-2
Sequence 611, App	31.9%	243	9	US-09-925-297-611
Sequence 53, Appl	11.8%	527	9	US-09-880-192-53
Sequence 53, Appl	11.8%	527	12	US-10-427-348-53
Sequence 42, Appl	9.2%	158	9	US-09-789-919-42
Sequence 12, Appl	9.1%	127	12	US-10-440-366-12
Sequence 2959, App	9.1%	127	12	US-10-144-047-2959
Sequence 5, Appl	7.2%	52	11	US-09-783-732-5
Sequence 6, Appl	6.2%	1965	12	US-10-359-012-6
Sequence 29, Appl	5.7%	1312	12	US-10-136-704-29
Sequence 29, Appl	5.7%	1312	15	US-10-071-179-29
Sequence 7, Appl	5.7%	2476	11	US-09-824-574-7
Sequence 8, Appl	5.6%	4723	12	US-10-359-012-8
Sequence 4, Appl	5.4%	1633	12	US-10-359-012-4

RESULT 1

US-09-783-732-4
; Sequence 4, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09783,732
; PRIOR FILING DATE: 2001-02-13
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-4

ALIGNMENTS

16	213	5.4	6642	12	US-10-369-493-5013	Sequence 5013, App
17	211	5.4	1480	12	US-10-359-012-22	Sequence 22, Appl
18	211	5.4	1790	12	US-10-369-493-1586	Sequence 1586, App
19	211	5.4	1961	12	US-10-028-248A-103	Sequence 103, App
20	210.5	5.4	200	10	US-09-928-412-2	Sequence 2, Appl
21	210	5.3	172	15	US-10-378-173-84	Sequence 84, Appl
22	209	5.3	1959	12	US-10-028-248A-36	Sequence 36, Appl
23	206	5.2	1961	12	US-10-028-248A-105	Sequence 105, Appl
24	206	5.2	3002	12	US-10-359-012-10	Sequence 10, Appl
25	206	5.2	3002	12	US-10-359-012-20	Sequence 20, Appl
26	204.5	5.2	1091	12	US-10-369-493-6328	Sequence 6328, App
27	203.5	5.2	1398	12	US-10-094-466-30	Sequence 30, Appl
28	203.5	5.2	1400	10	US-09-764-176-7	Sequence 7, Appl
29	203	5.2	1095	15	US-10-128-714-8305	Sequence 8305, App
30	200.5	5.1	1960	12	US-10-236-031B-62	Sequence 62, Appl
31	200.5	5.1	1960	12	US-10-028-248A-104	Sequence 104, App
32	196	5.0	2665	9	US-09-864-761-34248	Sequence 34248, A
33	196	5.0	3664	15	US-10-177-293-423	Sequence 423, App
34	195	5.0	205	15	US-10-395-403-116	Sequence 116, App
35	195	5.0	1234	12	US-10-203-311A-6	Sequence 6, Appl
36	194.5	5.0	1276	15	US-10-029-115-6	Sequence 32, Appl
37	194	4.9	2843	8	US-08-681-219-32	Sequence 30, Appl
38	194	4.9	2843	12	US-10-092-138-30	Sequence 30, Appl
39	194	4.9	2843	12	US-09-230-111C-30	Sequence 17, Appl
40	193.5	4.9	1501	10	US-09-924-154-17	Sequence 1, Appl
41	193	4.9	2843	10	US-09-987-482-1	Sequence 16, Appl
42	192.5	4.9	1616	11	US-09-820-843A-16	Sequence 3, Appl
43	192	4.9	1781	9	US-09-738-877-3	Sequence 13, Appl
44	192	4.9	1781	11	US-09-961-403-13	Sequence 7646, App
45	191.5	4.9	1881	12	US-10-032-585-7646	

181 TDAGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRASGRKUSSENSYSLDLEIGPGQ 240
181 TDAGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRASGRKUSSENSYSLDLEIGPGQ 240
241 LSSSTFSEKNESSRNLELPLSETSIKORMAKYQAAVSKQSSSTNTYNELKASGGEIKI 300
241 LSSSTFSEKNESSRNLELPLSETSIKORMAKYQAAVSKQSSSTNTYNELKASGGEIKI 300
301 HKMEQKENVPPGPEVCITHQGEKISANENSILAVRSTPAEDSDSDSVKSEVQPVHPKP 360
301 HKMEQKENVPPGPEVCITHQGEKISANENSILAVRSTPAEDSDSDSVKSEVQPVHPKP 360
361 LSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPVMERLLANQQVFHISCFRCSCY 420
361 LSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPVMERLLANQQVFHISCFRCSCY 420
421 NNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNEEILERPAP 480
421 NNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNEEILERPAP 480
481 LANARETPHSPGVEDAPIAKVGVLAASWEAKASSQOQEKEDKPAETKKLRIAWPPPTLGS 540
481 LANARETPHSPGVEDAPIAKVGVLAASWEAKASSQOQEKEDKPAETKKLRIAWPPPTLGS 540
541 SGSALEEGIKMSKPKWPPEDISKEPEVPEDVDLKLKRRSSSLKRSRPTTVAASFQST 600
541 SGSALEEGIKMSKPKWPPEDISKEPEVPEDVDLKLKRRSSSLKRSRPTTVAASFQST 600
601 SVKSPKTVSPPIRGWMSQSESVGGRVAERQVENAKSKNGNVGKTTWQNKESKG 660
601 SVKSPKTVSPPIRGWMSQSESVGGRVAERQVENAKSKNGNVGKTTWQNKESKG 660
661 ETGRKSKEGHSLMENENLVENGADSDDEDDNSFLKQSQPOEPKSLNWSVFDNTFAEFT 720
661 ETGRKSKEGHSLMENENLVENGADSDDEDDNSFLKQSQPOEPKSLNWSVFDNTFAEFT 720
721 TQNKSQDVELWEVVKELSVESQIKRNRYYDEDEDEE 759
717 TQNKSQDVELWEVVKELSVESQIKRNRYYDEDEDEE 755

RESULT 2
US-09-783-732-2
; Sequence 2, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-2

Query Match 77.1%; Score 3027.5; DB 11; Length 596;
Best Local Similarity 99.0%; Pred. No. 2.4e-194;
Matches 594; Conservative 0; Mismatches 1; Indels 5; Gaps 5;
161 MENCGLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRAS 220
1 MENCGLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRAS 60

221 GRKISENSYSLDLEIGPQQLSSSTFDSKNESSRNLELPLRSETSIKORMAKYQAAVSK 280
61 GRKISENSYSLDLEIGPQQLSSSTFDSKNESSRNLELPLRSETSIKORMAKYQAAVSK 120
281 QSSSTNTYNELKASGGEIKIHKMEQKENVPPGPEVCITHQGEKISANENSILAVRSTPAE 340
121 QSSSTNTYNELKASGGEIKIHKMEQKENVPPGPEVCITHQGEKISANENSILAVRSTPAE 180
341 DDS-RDSQVKSQVQPVHPKPLSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVP 399
181 DSDPGDSQVKSQVQPVHPKPLSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVP 240
400 MERLLANQQVFHISCFRCSCYNNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHR 459
241 MERLLANQQVFHISCFRCSCYNNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHR 300
460 PHKDLWASKNEEILERPAPLANARETPHSPGVEDAPIAKVGVLAASWEAKASSQOQEK 519
301 PHKDLWASKNEEILERPAPLANARETPHSPGVEDAPIAKVGVLAASWEAKASSQOQEK 360
520 DXPAETKKLRIAWPPPTLGSSESSALEEGIKMSKPKWPPEDISKEPEVPEDVDLKLK 579
361 DXPAETKKLRIAWPPPTLGSSESSALEEGIKMSKPKWPPEDISKEPEVPEDVDLKLK 420
580 RSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIRGWMSQSESVGGRVAERQVENA 639
421 RSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIRGWMSQSESVGGRVAERQVENA 480
640 KASKNGNVGKTTWQNKESKGKGTGRKSKEGHSLMENENLVENGADSDDEDDNSFLKQSQP 699
481 KASKNGNVGKTTWQNKESKGKGTGRKSKEGHSLMENENLVENGADSDDEDDNSFLKQSQP 537
700 QEPKSLNWSVFDNTFAEFTTQNKSQDVELWEVVKELSVESQIKRNRYYDEDEDEE 759
538 QEPKSLNWSVFDNTFAEFTTQNKSQDVELWEVVKELSVESQIKRNRYYDEDEDEE 596

RESULT 3
US-09-925-297-611
; Sequence 611, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (237)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (238)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (243)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-611

Query Match 31.9%; Score 1254; DB 9; Length 243;

Best Local Similarity 98.8%; Pred. No. 3.5e-76;
Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 321 EGEKISANENSLAVRSTPAEDDSRDQVSEVQVPHKPLSPDSRASSLSSESSPPKAMK 380
DB 1 EGEKISANENSLAVRSTPAEDDSRDQVSEVQVPHKPLSPDSRASSLSSESSPPKAMK 60
QY 381 KFOAPARETCVECKTQVPMERLLANQVPHISCFRCSYCNKKLSLGTVASLHGRYICKP 440
DB 61 KFOAPARETCVECKTQVPMERLLANQVPHISCFRCSYCNKKLSLGTVASLHGRYICKP 120
QY 441 HFNLFKSKGNYDEGFGHRRPKDOLWASKNENEBILERPQAQANARETPHSPGVEDAPIAK 500
DB 121 HFNLFKSKGNYDEGFGHRRPKDOLWASKNENEBILERPQAQANARETPHSPGVEDAPIAK 180
QY 501 VGVLAASMEAKASQOQEKEDKPAETKKLRIANPPPTLSSGSSALBEGIKMSKPKWPPED 560
DB 181 VGVLAASMEAKASQOQEKEDKPAETKKLRIANPPPTLSSGSSALBEGIKMSKPKWXXED 240
QY 561 E 561
DB 241 E 241

RESULT 4

US-09-880-192-53
; Sequence 53, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 058201CD1
US-09-880-192-53

Query Match 11.8%; Score 464.5; DB 9; Length 527;
Best Local Similarity 35.7%; Pred. No. 6.9e-23;
Matches 107; Conservative 47; Mismatches 105; Indels 41; Gaps 8;

QY 400 MERLLANQVPHISCFRCSYCNKKLSLGTVASLHGRYICKPHFNOLFKSKGNYDEGFGHR 459
DB 1 MECLVADKQNFHKSFCRCHCNKSLNYSALHQLYCKPHFKQLFKSKGNYDEGFGHK 60
QY 460 PHKDLWASKNENEEI-----LERPAQANARETPHSPGVEDAPIAKVGVLAASMEAKASQ 515
DB 61 QHDKRWCKNQSRVDFIPNEEPNCKNTAENTLVPGDNEHL-----DAGNSE 109
QY 516 QEKED--KPAETKKLRIANPPPTLSSGSSALBEGIKMSKPKWPE-DEISKPEVPEDVD 572
DB 110 GORNDLRLKGLGERGLKVIWPPSKPIKTLPFEEELKMSKPKWPEMTLLSPFKSESL 169
QY 573 LDLLKLRSSSLKERSRPFVAASFQSTSVKSPKTVSPPIRKGWSMSEQSESVGGRVAE 632
DB 170 LEDVTPENKQRODHFFP-LQPYLQSTHVC-----QKEDVIG--IKE 209
QY 633 RKQVENAKSKKNGVGTWQNKESKGETKRSKEGHSLEMEENLVENGADSDDEDNS 692
DB 210 MKMPEGRKDEKKE---GRKVVQDRPSEADTKSNRKSAMDNDNNNVIVQSAEKEKNEKT 266

RESULT 5

US-10-427-348-53
; Sequence 53, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 058201CD1
US-10-427-348-53

Query Match 11.8%; Score 464.5; DB 12; Length 527;
Best Local Similarity 35.7%; Pred. No. 6.9e-23;
Matches 107; Conservative 47; Mismatches 105; Indels 41; Gaps 8;

QY 400 MERLLANQVPHISCFRCSYCNKKLSLGTVASLHGRYICKPHFNOLFKSKGNYDEGFGHR 459
DB 1 MECLVADKQNFHKSFCRCHCNKSLNYSALHQLYCKPHFKQLFKSKGNYDEGFGHK 60
QY 460 PHKDLWASKNENEEI-----LERPAQANARETPHSPGVEDAPIAKVGVLAASMEAKASQ 515
DB 61 QHDKRWCKNQSRVDFIPNEEPNCKNTAENTLVPGDNEHL-----DAGNSE 109
QY 516 QEKED--KPAETKKLRIANPPPTLSSGSSALBEGIKMSKPKWPE-DEISKPEVPEDVD 572
DB 110 GORNDLRLKGLGERGLKVIWPPSKPIKTLPFEEELKMSKPKWPEMTLLSPFKSESL 169
QY 573 LDLLKLRSSSLKERSRPFVAASFQSTSVKSPKTVSPPIRKGWSMSEQSESVGGRVAE 632
DB 170 LEDVTPENKQRODHFFP-LQPYLQSTHVC-----QKEDVIG--IKE 209
QY 633 RKQVENAKSKKNGVGTWQNKESKGETKRSKEGHSLEMEENLVENGADSDDEDNS 692
DB 210 MKMPEGRKDEKKE---GRKVVQDRPSEADTKSNRKSAMDNDNNNVIVQSAEKEKNEKT 266

RESULT 6

US-09-789-919-42
; Sequence 42, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Kateri
; APPLICANT: Lemischka, Ihor
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-42


```

Query Match      9.2%; Score 363; DB 9; Length 158;
Best Local Similarity 58.7%; Pred. No. 8.4e-17;
Matches 64; Conservative 16; Mismatches 27; Indels 2; Gaps 1;

QY 362 SPDSRASSLSSESPPKAMKPF--QAPARETCVCQKTYPMERLLANQQVHFHISCFCSY 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 TPSHEAKGSGSGSTVQRSKSPFLPAQVKETCAACQKTYPMERLVADKLFIHNSCFCKH 70
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 420 CNNKLSLGTATSLHGRITYCPHFNFQLPKSKGNVDEGFGRHPKDLWASK 468
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 CHTKLSLGSVAAMHGEFYCPHFHFOQLPKSKGNVDEGFGRKHOKHRELWAHK 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-10-440-366-12
; Sequence 12, Application US/10440366
; Publication No. US20030181706A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Voikmuth, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
; FILE REFERENCES: EB-0006-1 CIP
; CURRENT APPLICATION NUMBER: US/10/440,366
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 09/855,323
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/195,292
; PRIOR FILING DATE: 1998-11-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 402234CD1
US-10-440-366-12

```

Query Match	9.1%	Score 357;	DB 12;	Length 127;
Best Local Similarity	50.7%;	Pred. No. 1.6e-16;		
Matches	69;	Conservative 16;	Mismatches 27;	Indels 24;
Gaps	3;			

QY	333	AVRSTAEEDSRSSQVYKSEVQPVHPKPLSPDGRASSLSSESPPKAMKTKQAPARETCVE	392
Db	7	AAQATPSHDAKGGG--SSTVQ-----RSKSFS-----LRAQVKETCAA	42
QY	393	CQTVVPMERLLANQQVHFHSICFRCSYCNKNKLSLGTVASHLGRLYCKPHENQIFKSKGNY	452
Db	43	CQTVVPMERLVDADKLIHFNSCFCCHCKHTKLSLGSYAAUHGFEYCKPHFQQLFKSKGNY	102
QY	453	DEGFGRHPKDLWASK	468
Db	103	DEGFGRKHOKELWAHK	118

```

RESULT 8
US-10-104-047-2959
; Sequence 2959, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392a1el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2959
; LENGTH: 127
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
US-10-104-047-2959

Query Match.
Best Local Similarity 50.7%; Score 357; DB 12; Length 127;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3

Qy 333 AVRSTPAEDDSROQVSEYQQVPHKPLSPDPSASSLSSESPKAMKKFOAPARETCVE 392
Db 7 AAQATPSHDAKGG--SSTVQ-----RSKSFSLDRAQVKETCAA 42

Qy 393 CQKTVYPMERLLANQQVHFHISCFRCSCYCNKLSLGTYSALHGRYCKPHFQKFSKGN 452
Db 43 CQKTVYPMERLIVADKLIFHNSCCFCKCHTKLSLSGSVAALHGFEYCKPHFQKFSKGN 102

Qy 453 DEGFGRHPRHKLWASK 468
Db 103 DEGFGRQHKELWAHK 118

```

```

RESULT 9
US-09-783-732-5
; Sequence 5, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (0)...(0)
; OTHER INFORMATION: mutant sterol regulatory element binding protein 2
US-09-783-732-5

```

Query Match 7.2%; Score 281; DB 11; Length 52;
Best Local Similarity 96.3%; Pred. NO. 5.7e-12;
Matches 50; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 390 CVCEQCTVTPMERLLANQOVFHSICRFSYCNKLSLGTVASLHGRITYKPH 441

Db 1 CVGCKTVPMERLLANQOVFHSICRFSYCNKLSLGTVASLHGRITYKPH 52

RESULT 10
 US-10-359-012-6
 ; Sequence 6, Application US/10359012
 ; Publication No. US20030232419A1
 ; GENERAL INFORMATION:
 ; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
 ; APPLICANT: KOLODKIN, Alex R.
 ; APPLICANT: TERMAN, Jon R.
 ; APPLICANT: MAO, Tianyi
 ; APPLICANT: PASTERKAMP, Ronald J.
 ; APPLICANT: YU, Hung-Hsiang
 ; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES,
 ; TITLE OF INVENTION: AND METHODS OF USING THE SAME
 ; FILE REFERENCE: JHU1840-3
 ; CURRENT APPLICATION NUMBER: US/10/359,012
 ; CURRENT FILING DATE: 2003-02-04
 ; PRIOR APPLICATION NUMBER: US 60/388,325

```

; PRIORITY FILING DATE: 2002-06-13
; PRIORITY APPLICATION NUMBER: US 60/384,302
; PRIORITY FILING DATE: 2002-05-30
; PRIORITY APPLICATION NUMBER: US 60/354,178
; PRIORITY FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 1965
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-359-012-6

Query Match
Best Local Similarity 21.7%; Score 245; DB 12; Length 1965;
Matches 166; Conservative 105; Mismatches 303; Indels 192; Gaps 35;

QY 54 RNTENLSQHFKGLTVLVKKWNPGLGAESHTSLRNSSTEIRHRAD-----HPPA 106
Db 450 QTTPEVNSKFNQSYSDPVR---VPNI---NVNFLRPS---QVRHLYDTGTEDIHLEM 500
QY 107 EYVSHAASAKAQDEQIHPRSRLSPPEALVQ-----RY 142
Db 501 ESLVNSRTTPKLTNRNESVARSKLLGWCQRTDGYAGNVVDTLTMWSKGLCALCAIHRY 560
QY 143 -PHIKDGEDLKDHSKWKWNCNLCESRHEVEKSEISENTDAS--GKIEKYNVP--LNRL 197
Db 561 RPDLDIFSLEQNVVE-KNNQAFADIAEKELGISPMTGKENVASVGEPPDKLSVMYLTQF 619
QY 198 KMFPEKGEPT-----QTKILRAQRS-----ASGRKISENSYSLODLEIGP 238
Db 620 YEMFKDSLPSDDTLNAAEKAVALIASTRSPISFLSKLQTIISRKSPDKKEXDLD-GA 678
QY 239 GOLSSSTFDSKNEBRRNL--ELPLSTSTIKDRMAKYQAAVSKQSSSTNYTNELKASGG 296
Db 679 GK-RRKTSQSEEEAPRGRGERPLVSTLTDRM---DVAVGNQNKVKYMATQLLA--- 731
QY 297 EIKIHMKQKENVPPGPEVCITHQSEKISANENSLAVRSTPAEDDSDSQVSEVQPV 356
Db 732 -----KFEEN---APAQSIGIRRQQTQERGASQPSCCCL---PGQ-----V 765
QY 357 HPKPLSPDSRASSLSSESPKAMKKFAPAPARETCVCKTVPYPMERLLANOQVPHISCFR 416
Db 766 RPAP-TPRWKQSGMKKEFPQ-----NLGSDTCYFCOKRYVYMERLSAEGKFHRSCKFK 818
QY 417 CSYCNKLSLGTYA--SHGRIYCKPFPNQLFKSGNVDGFGHRPHKDLWASKNENEI 474
Db 819 CEYCATTURLGAYAYDIEDGKFKYCKPHYCYRL-----SCYAQR-----856
QY 475 LERPAQL---ANARETPHSPGVEDAPIAKVGLVLAASMEAKASSOQEKDPAETKKLRIA 531
Db 857 -KRPAAVAPLSGKEAKGPLQDQATTANGRANAVASSTERTPGSGVNGLEEPSIAKRLR-G 914
QY 532 WPPTELGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPF 591
Db 915 TPERIELENYLSRQAEALO-----EVPEETQAE-----HNLSSVLDTQAAE 957
QY 592 TVAASFQSTSVKSPKTSVPPIRKGMSEQSEESVGGGRVAERKOVENAKASKKNGVNGKT 651
Db 958 DVASRSARRAAGRPPATFP-----EESSEAGNQL--QVVMHAADPLETQADVHWT 1006
QY 652 TWQNKESGKGTGKSKESCHSLEMENENLVNGADSDDDNSFLKQSQPQPFKSLNWSFV 711
Db 1007 HIREEEERMAPASESSAS-----GAPLDEND---LEEDVDSEPAETEGAAE 1052
QY 712 D-----NTFAE-----EFTTQNKSDVEL-----WEGEVVKELSVEE 744
Db 1053 DGDGDTGAELDDQHWSDSPSDADRELRLPCPAEGEAELELRVSE 1098

RESULT 11
US-10-126-704-29
; Sequence 29, Application US/10126704
```

```

; Publication No. US20030170647A1
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: 44. US5.DIV
; CURRENT APPLICATION NUMBER: US/10/126,704
; PRIORITY FILING DATE: 2002-04-20
; PRIORITY APPLICATION NUMBER: US 60/091,315
; PRIORITY FILING DATE: 1998-06-30
; PRIORITY APPLICATION NUMBER: US 60/111,909
; PRIORITY FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.PM
; SEQ ID NO 29
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 294..296
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 432..434
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 755..757
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 856..858
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 859..861
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 910..912
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1151..1153
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1226..1228
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 102..105
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 663..666
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 808..811
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 885..888
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 17..19
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 31..33
```

```

/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 41..43
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 100..102
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 216..218
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 471..473
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 507..509
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 531..533
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 591..593
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 656..658
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 801..803
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 812..814
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 815..817
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 876..878
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 888..890
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 939..941
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1060..1062
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1128..1130
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1129..1131
/ OTHER INFORMATION: potential

```

```

/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1135..1137
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1181..1183
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1208..1210
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1249..1251
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 47..50
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 126..129
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 157..160
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 158..161
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 159..162
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 216..219
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 274..277
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 276..279
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 295..298
/ OTHER INFORMATION: potential
/ FEATURE:

```

Query Match 5.7% Score 223; DB 12; Length 1312;
 Best Local Similarity 19.4% Pred. No. 3.5e-06;

Matches 177; Conservative 150; Mismatches 329; Indels 258; Gaps 42;

```

Qy 40 KYQKAABETNMKRSNTENLSQHFRTGLTLVLKKNWPNGLGAESHTDSLRSNSTEIRH 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 EFQWALPEKVNVKQCEENVKE-----IKVKEENETEIKEIKMBEERN-----IIP 449

Qy 100 RADHPAEVTSHAASGAKADQEEQIHP-----RSRLRSPPEALVQGRYPHIK---DGEDL 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 REEKPIED-----EIERKENIKPSLGSKKNLLESIPTHSDQKEVNIKKPEDNENL 500

Qy 152 KDHSTESKKWENCL-----GSRHEVEKSEISENTDAS 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 DDKDDDTTRVDESINIKVABEBAKSGDETKEDEDDDEAESEEEEEDEDDDDN 560

Qy 185 GKIEKYNVPLNLKWMFEKGEPTQTKILRAQSRAS-----GRKISENSYSL 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 NEEEFECYPPGMKVQVRYGRGKNQWYEASIKDSVGEVGLVLYVHYCGWNVRYDEWIK 620

```


NAME/KEY: PHOSPHORYLATION
LOCATION: 471..473
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 507..509
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 531..533
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 591..593
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 656..658
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 801..803
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 812..814
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 815..817
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 876..878
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 888..890
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1060..1062
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1128..1130
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1129..1131
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1135..1137
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1181..1183
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1208..1210
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1249..1251
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION

LOCATION: 47..50
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 126..129
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 157..160
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 158..161
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 159..162
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..219
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 274..277
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 276..279
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 295..298

Query Match

Best Local Similarity 5.7%; Score 223; DB 15; Length 1312;

Matches 177; Conservative 150; Mismatches 329; Indels 258; Gaps 42;

QY	40	KYKAAEETNMKKRSNTENLSQHFRKGTTLVLKKNENPGLGAESHTDLSRNSSTEIRH	99
DB	403	EFQMALPEKVNKQCKECENYKE-----IKVEENETEIKEMKEERN-----IIP	449
QY	100	RADHPPAEVTSAAAGAKADQEEQTHP-----RSRLRSPPEALVQGRYPHK---DGEDL	151
DB	450	REEKPIED-----EIEKENIKPSLGSKKKLLLESIPTHSQEKEVNIKKPEDNENL	500
QY	152	KDHSTESKMNCL-----GESRHEVEKSEISENTDAS	184
DB	501	DDKDDTTTRVDESINIKVEAEAEKAKSGDETCKEEDDEDEAEAEDEDEDEDEDDDN	560
QY	185	GKIEKYNVPLNRLXMMPEKGEPTQTKILRAQSRAS-----GRKISENSYSL	231
DB	561	NEEEFECYPPGMKVQVRYGRGKQKYEASIKOSDVEGGEVLVLVHYCGWNVRYDEWIK	620
QY	232	DDLEIGPGQLS-----SSTFDSEKNE-----SRNLELPLRLSETSIK-----DRMA	272
DB	621	ADKIVRPADKVVPKIKRKKIKNKLDKEKDEKYSRNCKLRSLKPPFPOTNPSPKWS	680
QY	273	KYQAAVSKQSSS-----TNYTNELKASGGEIKIHKMEKENVPPGPEVCITHQGEK	324
DB	681	KLDLTDKNSDTAHIKSIEITSILNGIQAESSADSEDEQEDRGAQDMNNGKESKIDH	740
QY	325	ISANENSLAVRSTPAEDDSRDSQVKSE-----VOQPVHPKP-----LSPDS--	365
DB	741	LNNRNDLI-----SKEQNSSSLLEENKVHADLVISKVSKSPELRKDKIEVLSEDTY	795
QY	366	-----RASSLSESSPPKAMK-KFOAPARETC-VECOKTVYP--MERLLANQOVPHIS	413
DB	796	BEDEVTKRGKVDKDDTKSKPOIKGKGRYCNTEELKTGSGKKEEAKNKE-----	850
QY	414	CFRCSYC-----NNKLSLGTYSALHGRIVCKPHEN	443
DB	851	-----SLCWENSSNSSDDEEETKAKMTPTKYNGLLEEKSLRTTGFYSG-----FS	899

QY 444 QL-----FKSGNYDEGFH---RPHKDLWAS-----KNENEILERPAQLANARETPHS 490
DB 900 EVAEKRIKLNNSDERLONSRADKRDVWSSIQOWPKTKLKELFS-DSDETAASPPH- 957
QY 491 PVEDA-----PIAKVGVLAASMEAKASQOEKEDKPAETKKLRI-----AWP----- 533
DB 958 PAPEEXVAEESXQVAEEESCSVELEKPPPPVNVDSKPIEEKTVEVDRKAEPSSGSN 1017
QY 534 ----PPT--ELGSS-----GSALKEGKMSKPKWPEDEIS--KPEVPEDVLD----- 574
DB 1018 SVLNTPTTTPESSSVTVTSGSQOSSVTVSEPLAPNQEEVRSIKGETSTIEVDVAGE 1077
QY 575 LKLLRSSSLKERSRPTTAAASQSTSVKSGPKTVSPPIKRGWMSQESVESVGRVAERK 634
DB 1078 LQDLQSGN-----SSPAGFASVSSSSNQPBEPHP--EKACTGQKRVKDAQGGSSSK 1131
QY 635 QVENAKA-----SKNGNVGKTTWQNKESKGETKRSK-----EGHSLE--MENENL 679
DB 1132 QKSHKATVYNNKKKGKTNSDSEELSAGESITKQPVKSVSTGMKSHSTKSPARTQSP 1191
QY 680 VENGADSDSDNSFLKQSQPEKSLNWS---SFVDN-IFAEFTTQNKSQDVELWEGE 735
DB 1192 GKCGKNGKDPD--LKPSNRLPRVYKWSFQMSDLENMTSAERITILQEKLOEIRKXLYS 1249
QY 736 VKEL-SVEEQIKR 748
DB 1250 LKSEVASIDRRKR 1263

RESULT 13
US-09-824-574-7
; Sequence 7, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Järne, Olli A.
; TITLE OF INVENTION: ARI4, Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match 5.7%; Score 222; DB 11; Length 2476;
Best Local Similarity 20.6%; Pred. No. 9.5e-06;
Matches 180; Conservative 124; Mismatches 295; Indels 274; Gaps 41;

QY 12 TSLSLVATKELSLVNNKSSAIV--EISKY-----QKAAETNWEK- 53
DB 551 SSVKLVSSKDSRGNIKSVTAKRVKELFVLTVPVLSNPSPIKGVDCQEVQEKNGRKS 610
QY 54 ---RSN-----TENLSQHFRTGLTVLKKWENPGIAGHESDTSLNSTSEIRHRADHPP 105
DB 611 GVARESEKCRPEEISD--ENNVTIL-----LESDLRSPRVKTTPLRRQTESNP 660
QY 106 AEVTS-HAASGAKAOEEOIHF--RSRLRSPPEALVQGRYPH-IXDGED--LKDHSTESK 159
DB 661 AMSNDEESNGTMKEKQMSGPIRKOKENSADCATDNPKPKHVPKAKQPVIGQONSOD 720
QY 160 KNENCLGESRHEVEKSEISENTDAGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRSA 219
DB 721 EMLAVLKEASQMGHSS--SSDTD-----INEP-----QNMHKK-----TGKDD 757
QY 220 SGRKISENSYSDDLLEIGPGQLSSSTFSEKXESRNLELPLRSETSI--KORVAKYQAA 277

DB 758 NGKRRKNS-----TSGSDFDTKKGS-----TETSIISKKRQNY--- 793
QY 278 VSKQSSNTYNTELKASGEIKIHKEQ-----KENVPPGPEVCITHQE--GEKISAN--- 328
DB 794 ----SESSNYDEL-----EREIKTMSRIGAAKSVPEKKEEDSSEDEKQKVVNDNGH 844
QY 329 ENSLAVRSTPAEDDGRDSQV-----SEVQOPVHPKPLSPDRASSLSSESP 376
DB 845 ERAKTTQEGSSADDTGDEGRGGSCSAGSIEKVRSGVBFREMLCKPGVSSDGAERPS 904
QY 377 KAMKQFAPARETCVECOKTVYPMERLLANQOVFHISCFRCSYCNKLSLGTYSALHGR 436
DB 905 VKEENVNSPEDK--RVSKTKEKTKHLSROS----- 933
QY 437 YCKPHNQLFKSGNYDEGFHGRPHKDLWASKNENEILERPAQLANARETPHSPGVEDA 496
DB 934 ----RKGKGGSDGTDTRFPKKEQSDSESEGEK-----KQSRQRPGTGK-KKA 975
QY 497 PIKAVGVLAASMEAKASQOEKEDKPAETKKLRTAWPPTELGGSSGSALEEGIKMSK--- 553
DB 976 POLKGETTKRQEWSSSD-----CTERL-----PEEEIG-----PFSKGIKOSKTD 1019
QY 554 ----PKWPPDEISKPEVPEDVD-----LDLKLRRSSSLKERSRPFVAA 595
DB 1020 AGGEKKGKWKDKCKEKEELSDSDVKLPKGGDCSDSEDKTRNRVSLREKKR-FSLPA 1078
QY 596 SFQSTSVKPKTVPPIKRGWMSQESVESVGRV-----AERKQ 635
DB 1079 ----KSPG-----KRPECSSSDTEKSLKGCCDSTEKPKRIDLRERNSSKNT 1125
QY 636 VENAKSKNGNVGKTTWQNKESKGETKRSKESHSLEMENENL-----VENGA 684
DB 1126 KEVKSASSSDAEGSSSDNKQKQRTSAKKKTGNTKEKKNLSRATPKRKQVDITSSSS 1185
QY 685 DSDDDNSFLKQSQPEKPSLNWSFVDNTPAEFTTQNKSQDVELW----- 732
DB 1186 DIGDDQNSAGESSDEQKI---KPVTENLVLPSTGFCQSSGDEALSKSVPTVDDDDD 1242
QY 733 ---EGEVVVELSVEEQIKRNRVYDED--EDEE 759
DB 1243 DNDPENRIAKKMLLEE-IKANLSSDEDSGSDDE 1274

RESULT 14
US-10-359-012-8
; Sequence 8, Application US/10359012
; Publication No. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEP
; TITLE OF INVENTION: AND METHODS OF USING THE SAME
; FILE REFERENCE: JHU1840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/388,325
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/384,302
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/354,178
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4723
; TYPE: PRT
; ORGANISM: Drosophila
US-10-359-012-8

Query Match 5.6%; Score 221; DB 12; Length 4723;
Best Local Similarity 17.4%; Pred. No. 2,7e-05;
Matches 184; Conservative 137; Mismatches 352; Indels 386; Gaps 37;

QY 15 SLRVAKELSLVNKNSAIVEIPS-----KYKAAEETNMKKNSNTENLSQHFRKGT 68
DB 646 SLDLTELE-SRIWLNLYDQICDLPRGEIPHIKHPKMDFDLQRKYRINHQAQDFESK-- 702
QY 69 LTVLKKWENPGLGAESHTSLNSSTEIRHAD-----HPPAEVTHAASGAKA----- 118
DB 703 LLATPKAKGPMQDAVDIPITVORRSVLEERAKQRHEQLLNIGGAAGAAGVAGSG 762
QY 119 ----DQEQIHRP-----SRLSPPEALVQGRPHIKGDELKDHSTESKXNENCLGESRHE 171
DB 763 TGITQONDTPRSKKRRQVDTANIEERQQLQIEENRQERMSKRRQORC-----HQ 817
QY 172 VEKSEISENTASKIEKYNVPLNRLKXMFKEGPTQTKILRAQSASGRKISSENSYSL 231
DB 818 TQNFYKSLQLLAG-----KLLREGGE-----AGVAEDGTTP 849
QY 232 DDLIEIGPGLSSSTFDEKXNESRNL-----ELPRUSETSIKDRMAKO----- 275
DB 850 EDYSIFLYROQAFVENDRVNDELKLLFDRDRGDIPSLPRTADEQFSDRIKXMEQRM 309
QY 276 -----AAVSKQSSSTNYTNELKAS-----GGEIKIKHMEQXENVPPEVCV 317
DB 910 GRGLGCKXPKDLMRAIGKIDNDNVNREIEKIELSKTEIHGPKGREKVPKWSK----- 966
QY 318 THOGEKISANENS LAVRSTPAEDDSDQVSEVQPVH--PKPL----- 361
DB 967 -----EQFQARQHKM--SKPQRQDSREAEKFKDIDQITIRNLDKQLKEGHNLDVGERGN 1018
QY 362 -----SPDSRAS-----LSESSPPKAMKKFOAPARETCVEQ 394
DB 1019 KVASIAQOGKKDEANSDEKNAGSNATTNTNTVTPKSSKVALAFKQAAASEKCRFC 1078
QY 395 KTVYPMERILLANQOVHISCFRCSCYNNKLSLGTVA-----SLHGRYCKPHF----- 442
DB 1079 QTVYPMKTTVEGLVLRNCLKCHCHTNLRLUGGYAFDRDDQGRLYCTQHFLPKPLP 1138
QY 443 ---NOLFKS-----KGNVDEGFGHRPHKDL----- 464
DB 1139 QRTNARKASAAQAPASPAVPTAGSVPTAAATSEHMDTTPRDQVDLLOTSRANASADAM 1198
QY 465 -----WA-----SKVNEEILERAQLANARETHPSGVEDAPIAKVGVL 505
DB 1199 SDDEANVIDEHWSGRNFLPESNDSQSELSSDESDESEMEEADDSFFGAQTLQ 1258
QY 506 AS-----MEAKASSQOEKEDKPAETKULRIAWPP-----PT-- 536
DB 1259 ASDWIKQYCEDDSDDDFYDSEDDGDDTEGEEFKARELRROEVRLQPLPANLPDT 1318
QY 537 -----ELGSSGALBEGIKWSKPKWPPPEDEISKP----- 565
DB 1319 ETEKLLNVNDKNMADGSSLSGNSFESARQSPSTPLTPTRVEMEQLERDAPRKFSE 1378
QY 566 -----EVPEDVDLKK-----LRRSSSLKER-----SR 589
DB 1379 IEAISEKLYHMNMVQXNDLEVLAKENLVKSGILRLKTLKEKLAENAAIAAGQVTP 1438
QY 590 PTVAAASFQSTS---VKSPTVSP-----IRKGMWSMSESESV 626
DB 1439 PSATAPGLQPKSFKDEKFEKVSPPQVVEPKPKPVDFNLDELKPKRP-NPEERPKQL 1497
QY 627 GORVAERKOVENAKSKNGNKGKTTWONKESKGETG-----KRSKEGHSLEWENENLV-- 680
DB 1498 -PRPSLKPPQPKPGSKSTNVSRS--NSLKNASNGSPVKKAPVSNNSKMQIEGILT 1554
QY 681 -----ENGADSDDD-----NSFLKQ-----QSP 699
DB 1555 LRKIQSSSDQDQDMVDVEDVERKENELSKLKEIQASSFAGTMDHIKSLQTLWPTVSA 1614
QY 700 QEPKSLNWSFVDNTFAEFTTQNKSQDVELWEGEVVK 738

DB 1615 QAPPSMDLSKFPNOKQEKSTSTSNKNQVTLKQVNLAK 1653

RESULT 15
US-10-359-012-4
; Sequence 4; Application US/10359012
; Publication NO. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEP
; FILE REFERENCE: JHU1840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/388,325
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/384,302
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/354,178
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1633
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-359-012-4

Query Match 5.4%; Score 213; DB 12; Length 1633;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 166; Conservative 92; Mismatches 284; Indels 248; Gaps 34;

QY 46 EETNMKKRSNTENLSQHFRKGTIL--TVLKKWENPGLGAESHTDSDRNSTETIRHRA 103
DB 708 EGGNQNVKXSWANQLLAKFEESTENPSLMKQERVSGIG-----KPVLCSSSGPVHSCCP 763
QY 104 PPAEVTSHAASGAKADQEQIHPRSRLRSPPEALVQGRPHIKGDELKDHSTESKXNEN 163
DB 764 KPEATSPSPPLK-----RQFVSVVVG--HVL--RELKQVSAGSE----- 801
QY 164 CLGSRHEVEKSEISENTDASGKIEKY-NVPLNRLKXMFKEG-----EPTQTKILRAQSR 217
DB 802 CLSRPWRARAKSDIQ-----LGCTENFATLPSTPRAQALSGVLRLQOQVEEKILQRAQ 856
QY 218 SASGRK-----ISENSYSLDDLEIGPQQLSSSTFDEKXNESRNLRLPRUSETSIKDMA 272
DB 857 NLANREFHTKNIREKAAHLASM-FGHG-----DFPNK----- 888
QY 273 KYQAAVSKQSSSTNYTNELKASGGEIKIHKWEKENVPPGPEVCITHOGEKISANENSL 332
DB 889 ----LLSKGLSHTH-----PPSPSRPLSPOP----- 911
QY 333 AVKSTPAEDDSDRSQVSEVQQP-----VHPKPLSP----- 363
DB 912 AASSSPSTVDSA-SPARKEKSPGFHPSHRLTVHPQLTVGVKSSGIGAAAABVLNLY 970
QY 364 --DSRASSLSESPPKAMKK---FOAPARETCVEQKTVYPMERILLANQOVHISCFRCS 418
DB 971 MNDHRPKAQATSPDLESMRKSFPLNLSGSDTCYFKKRVYVMERLSAEGHFFHRECFRCS 1030
QY 419 YCNKLSLGTVA--SLHGRYCKPHFNQPKNGNVDGFGHPRHDKLWASKXNENBEILE 476
DB 1031 ICATTIRLAAYTDCDEGKCYCKPHF---IHCKTN-----SKQRKRAELKQOREEAT 1081
QY 477 RPAQLANARETHPSGVEDAPIAKVGVLAA-----SME 509
DB 1082 WQEQEAPRDTPTTE---SSCAVAAITLEGSPPVHFSLPVLHPLLGLMLLDWNDSIPESVH 1138

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 20:07:07 : Search time 7226 Seconds
(without alignments)
4297.039 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSPFNRQWTSLSRVTA.....LSVBEQIKENRYDEDEEE 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2885711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USFTO.spool/US09890549/runat_06012004_094752_19821/app.query.fasta_1.903
-DB=GenEmbl -OPMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09890549 @C@N 1.1 4617 @runat_06012004_094752_19821 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb.ba.*
2: gb.hts.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	3927	100.0	3655	9	AF198454	Homo sapi
2	3927	100.0	3664	6	AX086386	Sequence
3	3927	100.0	3664	9	HSM800164	Homo sapi
4	3899	99.3	2536	9	AK000372	Homo sapi
5	3640.5	92.7	3550	9	AF198455	Homo sapi
6	3417.5	87.0	2164	9	BC001247	Homo sapi
7	3416	87.0	3277	9	AK000335	Homo sapi
8	2970	75.6	3232	10	BC031490	Mus muscu
9	2967	75.6	2370	10	AF307845	Mus muscu
10	2710	69.0	3997	10	AF307844	Mus muscu
11	2679	68.2	1754	9	AF218025	Homo sapi
12	2465	62.8	2207	6	BD160117	Primer fo
13	2465	62.8	2207	9	AK023649	Homo sapi
14	2460	62.6	2793	9	AK096172	Homo sapi
15	2458	62.6	2783	6	AX017487	Sequence
16	2458	62.6	2783	6	BD135177	Human nuc
17	2449	62.4	3348	9	AF157325	Homo sapi
18	2376.5	60.5	2667	9	AK000057	Homo sapi
19	1749.5	44.6	115345	9	AC008147	Homo sapi
20	1519	38.7	4425	10	CGU22818	Cricetulus
21	1450	36.9	1021	6	AX012180	Sequence
22	1421	36.2	2121	9	BC010664	Homo sapi
23	1324	33.7	195301	2	AC134548	Mus muscu
24	1209	30.8	237661	2	AC114454	Rattus no
25	1128	28.7	2631	5	AF307846	Danio rer
26	1127	28.7	2678	5	BC047797	Danio rer
27	1103	28.1	777	11	G56637	SHGC-102035
28	881	22.4	641	6	AX011639	Sequence
29	657.5	16.7	2264	9	AK097519	Homo sapi
30	641	16.3	3287	9	AB097518	Macaca fa
31	617.5	15.7	5264	9	HSM803760	Homo sapi
32	615.5	15.7	4548	9	HSM803292	Homo sapi
33	614.5	15.6	4796	9	HSM803668	Homo sapi
34	614.5	15.6	5175	9	HSM803318	Homo sapi
35	614.5	15.6	5216	9	HSM803644	Homo sapi
36	613.5	15.6	4885	9	HSM803277	Homo sapi
37	612.5	15.6	4815	9	HSM803690	Homo sapi
38	609.5	15.5	4225	9	HSM803683	Homo sapi
39	608.5	15.5	2710	6	AX143335	Sequence
40	608.5	15.5	2710	6	AX143335	Sequence
41	606	15.4	413	6	AR270117	Sequence
42	602.5	15.3	6524	9	HSM804604	Homo sapi
43	596.5	15.2	3825	6	AX375615	Sequence
44	576	14.7	3516	4	AY246700	Sus scro
45	575.5	14.7	195516	2	AC135476	Rattus no

ALIGNMENTS

AF198454 3655 bp mRNA linear PRI 10-JAN-2000
LOCUS Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA,
DEFINITION complete cds.
ACCESSION AF198454
VERSION AF198454.1 GI:6685006
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE EPLIN, epithelial protein lost in neoplasm
JOURNAL Oncogene, 18 (54), 7838-7841 (1999)
MEDLINE 20087188
PUBMED 10618726
REFERENCE 2 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
FEATURES
source
1. 3655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
gene
1. 3655
/gene="EPLIN"
102..2381
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAF23755.1"
/db_xref="GI:6685007"
/translation="MESSPNRQWTSLSRVTAKELSLVNKNKSAIVEIFSKYOKA
AETNMEKSNSTENUSQHPKRGTLVLKKNWENPGLEABESHTSLRNSSTIRHRAD
HPEAVTSHAAGKADQEEQIHPRLRSPPEALVQGRYPHIKQGEDILKORHSTESK
MENGLGSRHEVEKSEISETDASGKIEKYNVPLNRLKWMFKGEPOTQKILRAOSRS
ASGRKTSNYSYLDLLEIGPGLSSSTFQSEKNESRNLELPLRSITSIKDRKXVQA
AVSKQSSNTYNELKASGGEIKIHKWEQKENVPPGECVITHQSEKISANENSLAV
RSTPAEDSDSQKSEVQFPKPSLRASSLSSESPPKAMKKFOAPARETCVE
CQKTYPMERLAKQVHFHISCPKSCNNKLSLGYASLHGRYICKPHNFQKFSKG
NYDEGFGRHKDLMAKKNNEIILERPQLANARETHPSGVEDAPIAKVGLAASM
EAKASQOEEDPAETKRLIAWPPTELGSSGSALKEGIMKSKPKWPEDEISKPE
VPDQVDLDLKLRRSSSLKERSRFTVAASFOSTSVKSPKTSPPPIRKWSNSEQEE
SVGRVAERQVENAKSKNGNKGKTTWQNKESGKETGRKSKESHEMENENLVEN
GADSDDEDNLFKQGPQEPKSLNWSFVDNTFAEEFTTQNKQSDQVELWEGEVVKEL
SVBEQIKRNPYYDEDEDEE"
polyA_signal
3632..3637
/gene="EPLIN"
BASE COUNT 1197 a 737 c 810 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3,34e-206 Length: 3655
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x AF198454 (1-3655)
QY 1 MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db 102 ATCGAATCATCTCCATTATATAGACGGCAATGACCTCACTATCATTTAGGCTAACAGCC 161
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 162 AAGTTCAGGACCTCGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTATTCCTCAAG 221

QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 222 TACCAGAAAGCAGCTGAAGAAACAAACATGGAGAGAGAGAGTAACACCCAAAATCTC 281
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 282 TCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTGTTAAAGAGAGTGGGAGAACCCAGG 341
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 342 CTGGGAGCAGAGCTCTCACACAGACTCTCTTACGGAAACAGCAGCAGTATAGGCACAGA 401
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaIleSerGlyValAlaLysAlaAspGln 120
Db 402 CGAGACCATCTCTCTGCTGAAGTACAGCCAGCGTCTCTTCTGGAGCCCAAGCTCAGCCAA 461
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 462 GAAGAACAATCCACCCAGATCTAGACTCAGCTCAGCTCCTCTGAAAGCCTCTGTTCAAGGT 521
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 522 CGATATCCCCACATCAAGACCGTGAGGATCTTAAAGACCCTCAACAGAAAGTAAAAAA 581
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 582 ATGGAAATTTGCTAGGAGAAATCCAGGATGAGTAGAGAAATCAGAAATCAGTGAATAAC 641
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 642 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGAGCTTAAGATGATG 701
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 702 TTGAGAAGGTGAACCAACTCAACTAAGATTCTCCGGGCCCAAGCCGAGTCAAGT 761
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 762 GGAAGAGAGATCTCTGAAAACAGCTATTCTTAGATGACCTGGAAATAGGCCAGGTGAG 821
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 822 TTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTACAGCAATCTCGAACTTCCA 881
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 882 CGCCTTCTCAGAAACCTCTATTAAGGATCGAATGGCCAAAGTACCAGGCAGCTGTGTCCAAA 941
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
Db 942 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAAT 1001
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 1002 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGTCTGATCACCCTCAG 1061
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1062 GAAGGGGAAAAGATTTCTGCAAAATGAGAATAGCTGCGAGTCCGTTCCACCCCTGCCGAA 1121
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1122 GATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCCTGTCCATCCCAAGCCA 1181
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1182 CTAAGTCAGATTCAGAGAGCCTCCAGCTTTCTTGTAAAGTTCTCTCTCCCAAGCAATGAAG 1241
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValThrProMet 400
Db 1242 AAGTTTCAGGCACCTGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTATTCCTCAAG 1301

QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1302 GAGCGTCTCTGGCCACACAGCAGTGTTCACATCAGCTGCTTCCGTTGCTCTATTGC 1361
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerDeuHisGlyArgIleTyrCysLysPro 440
Db 1362 AACAAACAACCTCAGTCTAGGAACATATGTCATCTTTACATGGAAGATCTATTGTAAGCCT 1421
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1422 CACTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGCTTTGGCCACAGACCA 1481
QY 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluLeuLeuGluArgProAlaGln 480
Db 1482 CACAGGATCTATGGGCAAGCAAAATGAAACGAAGAGATTTTGGAGACCCAGCCAG 1541
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1542 CTTGCCAAATGCAAGGGAGACCCCTCAGAGCCAGGGGTAGAGATGCCCTATTGCTAAG 1601
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 1602 GTGGGTGCTCGGTGCAAGTATGGAAGCCAGCCAGCCCTCTCAGCAGGAGGAAGAC 1661
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySer 540
Db 1662 AAGCCAGCTGAACCAAGAGCTGAGATCGCTGGCCACCCCTGAGCTGAACTTGAAGT 1721
QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProGluAsp 560
Db 1722 TCAGGAAGTCCCTTGGAGGAGGAGTCAAAATGTCAAGGCCCAATGGCTCTCTGAAGAC 1781
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db 1782 GAAATCAGCAAGCCGAAATCTCTGAGATGTCGATCTAGATCTGAAGAAGCTAAGACGA 1841
QY 581 SerSerSerLysLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1842 TCTTCTTCACTGAAGGAAAGAAAGCCGCCATTCACGTAGACGCTTCATTTCAAGCACC 1901
QY 601 SerValLysSerProLysThrValSerProPheArgLysGlyTyrPheSerMetSerGlu 620
Db 1902 TCTGTCAAGAGCCCAAAATCTGTCCCCACCTATCAGGAAAGCTGGAGCATGTCAAG 1961
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 1962 CAGAGTGAACAATCTGTGGGTGAAGAGTTGCAGAAAGAAACAAGTGGNAATGCGAAG 2021
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrPheGlnAsnLysGluSerLysGly 660
Db 2022 GCTTCTAAGAAAGATGGGAATGTGGGAAACCAACCTGGCAAAACAAGAACTTAAAGGA 2081
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2082 GAGACAGGGAAGAGAGTAAGGAAGGTCAAGTTTGGAGATGGAGATGAGATCTTGTA 2141
QY 681 GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2142 GAAATGGTGCACACTCCGATGAAGATGATACAGCTTCTTAAACAACAATCTCCACAA 2201
QY 701 GluProLysSerLeuAsnTyrPheSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2202 GAACCAAGTCTCTGAATGGTGGATTTGTAGACAACACCTTTGCTGAAGAAATCACT 2261
QY 721 ThrGluAsnGlnLysSerGlnAspValGluLeuTyrGluGluValValLysGluLeu 740
Db 2262 ACTCAGAATCAGAAATCCCAAGGATGTGGAACCTGGGAGGGAAGTGGTCAAGAGCTC 2321
QY 741 SerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2322 TCTGTGAAGAACAGATATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2378

RESULT 2

AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wienann, S.
Human dna sequences
Patent: WO 0112659-A 338 22-FEB-2001;
German Human Genome Project (DB)
Location/Qualifiers
1. 3664
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3.35e-206 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-890-549-4 (1-759) x AX086386 (1-3664)
QY 1 MetGluSerSerProPheAsnArgGlnTyrThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGAAATCATCTCCATTTAATAGACGCGCAATGGACCTCACTATCATTTAGGGTAAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTTCTCTGTCTCAACAGAACAGATCATCGGCTATTGTGGAATATTTCTCCAG 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTGAAGAAACAACATGAGAGAGAGAGAGTAACACCGAAATCTC 273
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTyrGluAsnProGly 80
Db 274 TCCAGCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGGAGAACCCAGGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACACAGCAGCACTGAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 394 CGAGACCAATCTCTCTGCTGAAGTGAAGCCACCGCTGCTCTGAGGCCAAAGCTGACCAA 453
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 454 GAAGAAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAGGCCCTCTGTTAGGGT 513
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 514 CGATATCCCCACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAGAAA 573
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 574 ATGGAATTTCTTAGAGAAATCCAGCATCAAGTATAGAAAAATCAGAAATCAGTGAAGAAC 633
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 634 ACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAACAGGCTTAAAGATGAT 693

QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuAsgAlaGlnSerArgSerAlaSer 220
DB 694 TTTGAGAAAGGTGACCAACTCAAACTAAGATTCTCCGGGCCCAAGCGGAGGTGCAAGT 753
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
DB 754 GGAAGGAAGATCTCTGAAAAAGAGTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAC 813
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgAsnLeuGluLeuPro 260
DB 814 TTGTCTATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGCAAAATCTGAAATCTCCA 873
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
DB 874 CGCTCTCAGAAACCTCTATAAGAGATCGAATGGCCAAAGTACCAGGAGCTGTGTCCAAA 933
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
DB 934 CAAGCAGCTCAACCAACTATACAAATCAGCTGAAAGCCAGTGGTGGCGAAATCAAAATT 993
QY 301 HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
DB 994 CATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCACCCTATCAG 1053
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1054 GAAGGGGAAAGATTCTGCAAAATGAGAAATAGCTGGCAGTCCGTTCACCCCTGCCGAA 1113
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1114 GATGACTCCCTGACTCCAGGTTAAGAGTCAAGTCAACAGCTGCTCCATCCCAAGCCA 1173
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProLysAlaMetLys 380
DB 1174 CTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTCTCCTCCCAAGCAATGAAG 1233
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1234 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTCAAGAGACAGTCTATCCAATG 1293
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1294 GAGCGTCTCTTGGCCCAACAGCAGGTGTTTCACATCAGCTGCTTCCGTGCTCTATTGC 1353
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1354 AACCAACAACTCAGTCTAGGAACATATGCTATCTTACATGGAAGATCTATTGTAGCCT 1413
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1414 CACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTTGGAGACCCAGCCAG 1473
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluGluIleLeuGluArgProAlaGln 480
DB 1474 CACAAGGATCTATGGGCAAGCAAAATGAAACCAAGAGATTTTGGAGACCCAGCCAGCCAG 1533
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1534 CTTGCAAAATGCAAGGAGAGACCCCTCAGAGCCAGGGTAGAGATGCCCTATTGCTAAG 1593
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
DB 1594 GTGGGTGTCTGGTGTGAAGTATGGAAGCCCAAGGCTCCTCTCAGCAGGAGAGGAAGAC 1653
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
DB 1654 AAGCCAGCTGAAACCAAGAGCTGAGATCGCTGGCCACCCCTCAGTGAACCTGGAAGT 1713
QY 541 SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProGluAsp 560
DB 1714 TCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAAGGCCCAAAATGGCCCTCTCTGAGAC 1773
QY 561 GluLysSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580

DB 1774 GAATCAGCAGCCCGAAGTCTCTGAGATGTCGATCTAGATCTGAAGAAGCTAAGACGA 1833
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1834 TCTTCTTCACTGAAGGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAAACGACC 1893
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSerGlu 620
DB 1894 TCTGTCAAGAGCCCAAAAATCTGTCTCCCACTATCAGGAAAGGCTGGAGCATGTGAGAG 1953
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
DB 1954 CAGAGTGAAGAGTCTCTGGGTGAAGAGTTGAGAAAGGAAACAAGTGGAAATGCCAAG 2013
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
DB 2014 GCTTCTAAGAGATGCGGAATGTGGGAAAAACAACCTGCGCAAAACAAGAAATCTAAGGA 2073
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
DB 2074 GAGACAGGAAGAGAGTAAAGAGGTGATAGTTTGGAGATGGAGATCAGAATCTTGTGA 2133
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnSerProGln 700
DB 2134 GAAATGGTGCAGACTCCCATGAAGATGATAACAGCTTCTCTCAAAACAATCTCCACAA 2193
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
DB 2194 GAACCCAGTCTCTGATTTGGTTCGAGTTTGTGAGCAACACCTTTGCTGAAGAAATTCAT 2253
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
DB 2254 ACTCAGAATCAGAAATCCAGGATGTGNACTCTGGAGGAGAGAGTGTCAAGAGCTC 2313
QY 741 SerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
DB 2314 TCTGTGGAAGACAGATAAAGAGAAATCGTATTATGATGAGATGAGGATCAAGAG 2370
RESULT 3
LOCUS HSM800164 3664 bp mRNA linear PRI 03-AUG-2001
DEFINITION Homo sapiens mRNA; cDNA DKFZp586i1918 (from clone DKFZp586i1918); complete cds.
ACCESSION AL136911
VERSION AL136911.1 GI:12053316
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3664)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2001) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp586i1918) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/
FEATURES
Location/Qualifiers
1..3664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="12"
/clone="DKFZp586i1918"
/tissue_type="uterus"

```
/clone_lib="586 (synonym: hutel). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1. 3664
/gene="DKFZp58611918"
94. 2373
/gene="DKFZp58611918"
/notes="similarity to transcription factor SF3"
/codon_start=1
/product="hypothesized protein"
/protein_id="CAB66845.1"
/db_xref="GI:12053317"
/db_xref="SPIREMBL:Q9UHB6"
/translation="MESSPNRRQWTSLSRVTAKELSLVNKNKSAIIVEIFSKYOKA
AEETNMEKRSNTENISOHERKGLTLVKKKWNPGLSAHSHTDSSTETIRHRAD
HPAEVTSAAAGAKAQEQEIHPRSRLRSPREALVQGRYPHIKDGEDLKDHSTESKK
MNCLEGRSEVSEKSEISENTDASGKIEKYNVPLNLRKMFKEGETQIKILRAQSR
ASGRKSENSYSJDDLEIGCGUSSSTFDSEKNESSRNJELPELSESTIKDRMAKYQA
AVSQSSSTYNELKASGEIKIKHWEQENVPPOPEVCIITHQEGEKISANENSLAV
RSTPAEDSDSVQKEVQVPHKPLSPDSRASISSESPPKAMKKFOAPARETCVE
CQTVVPMERLLANQVHFISCFRSCYNNKLISLGTVSLHGRIYCKPHFNOLFKSKG
NYDEFGHPRKDLWASKNEBILERPOLANARETTPHSPGVEDAPIAKVGLVLAASM
EAKASQOEDEDPAETKLRIRIAMPPTTELSSGSALEBGIKMSKRPWPEDEISKPE
VPEDVDJLKKLRSSLSKERSRPFVTAASFQTSVSKSPKTVSPPIRKGMSSEQSEE
SVGGRVAERQVENAKASKKNGVGTWQNKESKGTGKRSKEGHSLEMEENILVEN
SVEEQIKRNRYYDEDEDE"
polyA_signal 3637. 3642
/gene="DKFZp58611918"
3647
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3.35e-206 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x HSM800164 (1-3664)
QY 1 MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGAATCATCTCCATTTAATAGACGGCAATGACCTCACTATCATTTGAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTTCTCTGTCTACACAGAACAAAGTCAATCGGCTATTGTGGAAATATTTCTCCAA 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTCAAGAAACAAACATGAGAGAGAGAGAGAGTAACACCCGAAATCTC 273
QY 61 SerGluHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCACAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAAGTGGGAGAACCCAGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCACACAGACTCTTACCAGGACAGAGAGACTGAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValAlaLysAlaAspGln 120
Db 394 GCAGACCATCTCTCTGCTGAAGTACAGACGACGCTGCTTCTGGAGCAACCAAGCTGACCAA 453
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 454 GAAGAACAAATCCACCCAGACTAGACTCAGGTCACCTCTCTGAGAGCCCTCGTTGAGGGT 513
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
/clone_lib="586 (synonym: hutel). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1. 3664
/gene="DKFZp58611918"
94. 2373
/gene="DKFZp58611918"
/notes="similarity to transcription factor SF3"
/codon_start=1
/product="hypothesized protein"
/protein_id="CAB66845.1"
/db_xref="GI:12053317"
/db_xref="SPIREMBL:Q9UHB6"
/translation="MESSPNRRQWTSLSRVTAKELSLVNKNKSAIIVEIFSKYOKA
AEETNMEKRSNTENISOHERKGLTLVKKKWNPGLSAHSHTDSSTETIRHRAD
HPAEVTSAAAGAKAQEQEIHPRSRLRSPREALVQGRYPHIKDGEDLKDHSTESKK
MNCLEGRSEVSEKSEISENTDASGKIEKYNVPLNLRKMFKEGETQIKILRAQSR
ASGRKSENSYSJDDLEIGCGUSSSTFDSEKNESSRNJELPELSESTIKDRMAKYQA
AVSQSSSTYNELKASGEIKIKHWEQENVPPOPEVCIITHQEGEKISANENSLAV
RSTPAEDSDSVQKEVQVPHKPLSPDSRASISSESPPKAMKKFOAPARETCVE
CQTVVPMERLLANQVHFISCFRSCYNNKLISLGTVSLHGRIYCKPHFNOLFKSKG
NYDEFGHPRKDLWASKNEBILERPOLANARETTPHSPGVEDAPIAKVGLVLAASM
EAKASQOEDEDPAETKLRIRIAMPPTTELSSGSALEBGIKMSKRPWPEDEISKPE
VPEDVDJLKKLRSSLSKERSRPFVTAASFQTSVSKSPKTVSPPIRKGMSSEQSEE
SVGGRVAERQVENAKASKKNGVGTWQNKESKGTGKRSKEGHSLEMEENILVEN
SVEEQIKRNRYYDEDEDE"
polyA_signal 3637. 3642
/gene="DKFZp58611918"
3647
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3.35e-206 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x HSM800164 (1-3664)
QY 1 MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGAATCATCTCCATTTAATAGACGGCAATGACCTCACTATCATTTGAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTTCTCTGTCTACACAGAACAAAGTCAATCGGCTATTGTGGAAATATTTCTCCAA 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTCAAGAAACAAACATGAGAGAGAGAGAGTAACACCCGAAATCTC 273
QY 61 SerGluHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCACAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAAGTGGGAGAACCCAGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCACACAGACTCTTACCAGGACAGAGAGACTGAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValAlaLysAlaAspGln 120
Db 394 GCAGACCATCTCTCTGCTGAAGTACAGACGACGCTGCTTCTGGAGCAACCAAGCTGACCAA 453
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 454 GAAGAACAAATCCACCCAGACTAGACTCAGGTCACCTCTCTGAGAGCCCTCGTTGAGGGT 513
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
```

```
Db 514 CGATATCCCCACATCAAGGACGCTGAGGATCTTAAGACCATCTCAACAGAAAGTAAAAA 573
QY 161 MetGluAsnCysLeuGlyCysSerArgHisGluValGluLysSerGluLysSerGluAsn 180
Db 574 ATGGAAAAATTTCTAGGAGATCCAGGATGAAGTAGAAAAATCAGAAATCAGTGAAGAAC 633
QY 181 ThrAspAlaSerGlyLysLysLeuLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 634 ACAGATGCTTCGGGCAAAAATAGAAAAATATAATGTTTCCGCTCAACAGGCTTAAGATGATG 693
QY 201 PheGluLysGlyGluProThrGlnThrLysLysLeuArgAlaGlnSerArgSerAlaSer 220
Db 694 TTTGAGAAAGGTGAACCAACTCAAACTAGATTCTCCGGGCCCAAGCCGCAAGTCAAGT 753
QY 221 GlyArgLysLysSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 754 GGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAAAATAGGCCACAGT 813
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 814 TTGTCTATCTTCTACATTTGACTCGGAGAAAAATAGAGTAGACGAGAAATCTGGAACCTCCA 873
QY 261 ArgLeuSerGluThrSerLysLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 874 CGCTCTCAGAAAACTCTATAAAGGATCGAATGGCCAAAGTACCGAGCAGCTGTGTCCAAA 933
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysAlaSerGlyGlyGluLysLysLys 300
Db 934 CAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 993
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 994 CATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGCTCTGAGGCTGTGATCACCCTCAG 1053
QY 321 GluGlyGluLysLysLysSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1054 GAAGGGGAAAAAGATTTCTGCAAAATAGCAATAGCCCTGCGAGTCCGTTCCACCCCTGCCGAA 1113
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1114 GATGACTCCCTGACTCCAGGTTAAGAGTGAAGTCAACAGCCTGTCCATCCCAAGCCA 1173
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1174 CTAAGTCCAGATTCAGAGCCTCCAGCTTTCTGAAAGTTCTCTCCCAAGCAATGAAG 1233
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValtyrProMet 400
Db 1234 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTTCAGAGACAGCTCTATCCAAATG 1293
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisLysSerCysPheArgCysSerTyrCys 420
Db 1294 GAGCGTCTCTTGGCCCAACCCAGCAGGTGTTTCAATCAGCTGCTTCCGTTGCTCTTATTC 1353
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1354 AACACAAACTCAGCTTAGGAACATATGCAATCTTTATAGGAGAAATCTATTGTAAGCCT 1413
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1414 CACTTCAATCAACTCTTTAAATCTAAGGGCACTATATGATGAGGCTTTGGGACAGACCA 1473
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluLeuLeuGluArgProAlaGln 480
Db 1474 CACAAGGATCTATGGGCAAGCAAAAATGAAAAACGAGAGATTTTGGAGAGACCCAGCCAG 1533
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1534 CTTGCAAAATCAAGGGGAGACCCCTCAGCCAGGGGTAGAGATGCCCTATTGTTAAG 1593
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
```


385 GCAGACCATCTCTGCTGAAAGTGACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCAA 444
QY
121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB
445 GAAGAACAAATCACCACCCAGATCTAGACTCAGTCACTCTCTGAGCCCTCGTTTCAAGGCT 504
QY
141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB
505 CGATATCCCAACATCAAGAGCCGTGAGGATCTTAAAGACCATCTCAACAGAAAGTAAATAA 564
QY
161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluAsn 180
DB
565 ATGGAATAATGTCTAGGAATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAATAAC 624
QY
181 ThrAspAlaSerClyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB
625 ACAGATGCTCTGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATG 684
QY
201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB
685 TTTGAGAAAGGTGAACCACTCAACTAAGATTCTCCGGGCCCAAGCCGAATAGTCAAGT 744
QY
221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
DB
745 GGAAGCAAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAATAGCCCCAGTCAAG 804
QY
241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgAsnLeuGluLeuPro 260
DB
805 TTGTCATCTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAATCTGGAACCTTCCA 864
QY
261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
DB
865 CGCCTCTCAGAACCTCTCTATAAGAGTGAATGGCCAGTACGAGCAGCTGTGTCCAAA 924
QY
281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
DB
925 CAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAATCAAAAT 984
QY
301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB
985 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGTCTGCATCACCCTCAG 1044
QY
321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB
1045 GAAGGGGAAAGATTTCTGCATAATGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1104
QY
341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB
1105 GATGACTCCCGTGACTCCAGGTTAAGATGAGGTTCAACAGCCTGTCCATCCCAGCCA 1164
QY
361 LeuSerProAspSerArgAlaSerLeuSerGluSerSerProProLysAlaMetLys 380
DB
1165 CTAAGTCCAGATTCAGAGCCTCCAGTCTTCTGAAAGTTCTCTCCCAACCAATGAAG 1224
QY
381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB
1225 AAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTCAAGACAGCTCTATCCAATG 1284
QY
401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB
1285 GAGCGTCTTTGGCCCAACAGAGGTTTTCATCATCAGCTGCTCGTTCGTTGCTCTATTGC 1344
QY
421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB
1345 AACCAACAACCTAGCTCTAGACATATGATCTTTACATGGAGAAATCTATTGTAAAGCT 1404
QY
441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB
1405 CACTTCAATCAACTCTTTAAATCTAAGGCAACTATGATGAAGGCTTTTGGGCACAGACCA 1464
QY
461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
DB
1465 CACAAGGATCTATGGGCAAGCAAAAAATGAAGCAAGAGATTTTGGAGAGACAGCCCGAG 1524

QY 481 LeuAlaIleAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1525 CTTCGAAATGCAAGGGAGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTTATTGCTAAG 1584
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
DB 1585 GTGGGTCTCTGGCTGCAGATATGGAAGCCAGGCTCTCTCTCAGCAGGAGAAAGAGAC 1644
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
DB 1645 AAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGGCCACCCCTGAACTTGGAACT 1704
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTyrProProGluAsp 560
DB 1705 TCAGGAATGCTCTGGAGGAAGGATCAAAATGTCAAAGCCCAATGCCCTCTCTGAAAGAC 1764
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
DB 1765 GAATTCAGCAAGCCCGAAGTTCTGAGATGTCATCTAGATCTGAAGAGCTTAAGACGA 1824
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1825 TCTTCTTCTACAGGAAGAAAGAGCCGCCCTTACTGTAGCAGCTTCTATTCAAGCACC 1884
QY 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
DB 1885 TCTGTCAAGACCCCAAAACTGTGTCCCACTATCAGGAAGGCTGGAGCATGTCAAG 1944
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
DB 1945 CAGAGTGAAGAGTCTGTGGTGGAGAGATTGCAAGAAAGGAAACAAAGTGGAAAAATGCCAAG 2004
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
DB 2005 GCTTCTAAGAGAAATGGGAATGCGGAAAAACAACCTGGCCAAAAACAAGAAATCTAAAGGA 2064
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
DB 2065 GAGACAGGAAGAGAGATGAAGAGGTCAATAGTTTGGAGATGGAGAAATGAGAACTTTGTA 2124
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
DB 2125 GAAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAAAACAACA-TCTCCACAA 2183
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
DB 2184 GAACCCAAAGTCTCGAATTTGGTTCGAGTTTGTAGACCAACACCTTTGCTGAAAGAAATTTCACT 2213
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
DB 2244 ACTCAGAAATCAGAAATCCAGGATGTGAACCTTGGAGGGAGAGAGTGGTCAAAGAGCTC 2303
QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
DB 2304 TCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2360
RESULT 5
AF198455
LOCUS
DEFINITION Homo sapiens epithelial protein lost in neoplasm alpha (EPLIN)
ACCESSION AF198455
VERSION AF198455.1 GI:6685008
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 (bases 1 to 3550)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Maul, R.S. and Chang, D.D.
TITLE EPLIN, epithelial protein lost in neoplasm
JOURNAL Oncogene 18 (54), 7838-7841 (1999)

MEDLINE	20087188	Db	519	GAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTTCGCTGAAC	578
PUBMED	10618726	Qy	196	ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuArgAlaGln	215
REFERENCE	2 (bases 1 to 3550)	Db	579	AGGCTTAAGATGATGTTTCAGAAAGGTGAACCAACTCAAGATTCTCCGGGCCAA	638
AUTHORS	Maul,R.S. and Chang,D.D.	Qy	216	SerArgSerAlaSerGlyArgLysLysLeuSerGluAsnSerTyrSerLeuAspLeuGlu	235
TITLE	Direct Submission	Db	639	AGCCGAAGTCAAGTGAAGGAGATCTCTGAAACAGCTATTTCTAGATGACCTGGAA	698
JOURNAL	Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA	Qy	236	IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg	255
FEATURES	Location/Qualifiers	Db	699	ATAGGCCACAGGTCAGTTCATCTTCTACATTGACTCGAGAAATAGAGTAGACGA	758
source	1..3550	Qy	256	AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln	275
gene	/organism="Homo sapiens"	Db	759	AATCTGGAACCTCCACGCTCTCAGAAACCTCTATAAGGATCGAATGCGCAAGTACCAG	818
CDS	/mol_type="mRNA"	Qy	276	AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly	295
	/db_xref="taxon:9606"	Db	819	GCAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT	878
	/chromosome="12"	Qy	296	GlyGluLysLysLysLysMetGluGlnLysGluAsnValProProGlyProGluVal	315
	/map="12q13"	Db	879	GCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTCCCCAGGTCTCTGAGTTC	938
	1..3550	Qy	316	CysIleThrHisGlnGluGlyLysLysLeuSerAlaAsnGluAsnSerLeuAlaValArg	335
	/gene="EPLIN"	Db	939	TGATCACCATCAGAGAGGGGAAAGATTTCGCAATAGAGAAATGAGCTTGGCAGTCCGT	998
	/gene="EPLIN"	Qy	336	SerThrProAlaGluAspAspSer---ArgAspSerGlnValLysSerGluValGlnGln	354
	/codon_start=1	Db	999	TCCACCCCTCGCGAAGATGACTCCCGCAGGTGCTCCAGGTTAAGAGTGAAGTTCACAG	1058
	/product="epithelial protein lost in neoplasm alpha"	Qy	355	ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer	374
	/protein_id="AF23756.1"	Db	1059	CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCAGTCTTTCTGAAAGTTCT	1118
	/db_xref="GI:6685009"	Qy	375	ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln	394
	/translation="MENCLGESRHEVBKSEISENTDASGKIEKYNVPLNLKMFEGK	Db	1119	CTTCCCAAGCAATGAAGAAGTTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAG	1178
	EPTOTKILRAOSRASGRKISSENSYSLDDLEIGPOLSSSTFDSEKNESRNLEPL	Qy	395	LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCys	414
	SETSIKORMAXQAAVSKQSSSTNYTNELKASGGEIKHKWEQENVPVPGPEVCIHQ	Db	1179	AGACAGTCTATCCAATGGAGGCTCTCTTGGCCCAACCCAGCAGTGTTCACATCAGCTGC	1238
	EGKISANENSLAVRSPAEDDSFGDSQVSEVQVPHRPLSPDRSSALSSSSPPK	Qy	415	PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly	434
	AMKFPQARETCVEQTVYMERLLANQQVPHISCFRCSCNNKLSLGTYSLGR	Db	1239	TTCCGTTGCTCTATTGCAACCAACAACTCAGCTAGGAACATATGCACTCTTTACATGGA	1298
	IYCKPHNQLPFKSGNYDEGFGHRPHKDLWAKNENEIILERPAQLANARETPHSPGV	Qy	435	ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu	454
	EDAPIKGVLAASMEAKASQOEKDPKPAETKKLRIAMPPPTLGSAGSALBEGIKM	Db	1299	AGAATCTATTGTAAGGCTCCTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAA	1358
	SKPKWPPDETSKPEVEDVDLJKLRRSSSLKRRSPRTVAASFOSTSVKSPKTVS	Qy	455	GlyPheGlyHisArgProHisLysLeuTyrAlaSerLysAsnGluGluIle	474
	PPIPKGMSQSESQGRVAERQVENAKSKNGNVKKTWNKESKSGETGKRK	Db	1359	GGCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAAATGAAACCAAGAGATT	1418
	EGHLEMEENLVENGADDDNSFLKQSPQEPKSLNWSFFVDNTFABEFTTQNOK	Qy	475	LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGlu	494
	SQDVELMEVEVQKLSVEEQIKRNYRDEDEDEE	Db	1419	TTGGAGACACCCAGCCAGCTTGCNAATGCAAGGAGACCCCTCAGAGCCAGGGGTAGAA	1478
	3527..3532	Qy	495	AspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSer	514
	/gene="EPLIN"	Db	1479	GATGCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCCAAGCCCTCTCT	1538
		Qy	515	GlnGlnLysLysLysLysProAlaGluThrLysLysLeuArgLysLeuAlaTyrProPro	534
		Db	1539	CAGCAGAGAGAGAGACAGCCAGCTGAAACCAAGAGAGCTGAGATCGCTGGCCACCC	1598
		Qy	535	ProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysMetSerLysPro	554

polya_signal

BASE COUNT 1153 a 716 c 773 g 908 t
ORIGIN

Alignment Scores:

Pred. No.: 1,61e-190 Length: 3550
Score: 3640.50 Matches: 703
Percent Similarity: 99.72% Conservative: 0
Best Local Similarity: 99.72% Mismatches: 1
Query Match: 92.70% Indels: 1
DB: 9 Gaps: 1

US-09-890-549-4 (1-759) x AF198455 (1-3550)

Qy	56	AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys	75
Db	159	AACACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAAAG	218
Qy	76	TrpGluAsnProGlyLeuGluValGluSerHisThrAspSerLeuArgAsnSerThr	95
Db	219	TGGGAAACCCAGGCTCGGAGCAGAGTCTCACAGACTCTTACCGAAGCAGCAGCACT	278
Qy	96	GluIleArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly	115
Db	279	GAGATTAGCAGACAGACAGACCATCTCTCTGCTGAGTGACAAAGCCAGCTGCTTCTGGA	338
Qy	116	AlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGlu	135
Db	339	GCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAA	398
Qy	136	AlaLeuValGlnGlyArgTyrProHisLysLysAspGlyGluAspLeuLysAspHisSer	155
Db	399	GCCCTGCTTCAAGGTCGATATCCCCACATCAAGACCGTGAGGATCTTAAAGACACTCA	458
Qy	156	ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer	175
Db	459	ACAGAAAGTAAAAATCGAAAAATGCTTAGAGAAATCCAGGATGAAGTAGAAAAATCA	518
Qy	176	GluIleSerGluAsnThrAspAlaSerGlyLysLysLysLysLysLysLysLysLysLys	195

Db 1599 CCCACTGAATCTTGGAAAGTTTCAGGAAGTGCCTTGGAGGAAGGATCAAAAATGTCAAAGCCC 1658
Qy 555 LysTrpProGluAspGluLeuSerLysProGluValProGluAspValAspLeuAsp 574
Db 1659 AAATGGCCCTCCTGAAGACCAATCAGCAAGCCGGAAGTTCCTGAGGATGCTGATCTAGAT 1718
Qy 575 LeuLysLysLeuArgSerSerSerLeuLysGluArgSerArgProPheTrnValAla 594
Db 1719 CTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGGAGCGCCATCTACTGTAGCA 1778
Qy 595 AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProProleArgLys 614
Db 1779 GCATCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACGTGTGCCACCTATCAGAAA 1838
Qy 615 GlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAlaGluArgLys 634
Db 1839 GCGTGGAGCATGTCAGAGCAGAGTGAAGATCTGTGGTGGAGAGTTCAGAAAAGAAA 1898
Qy 635 GlnValGluAsnAlaLysLysLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1899 CAAGTGGAAAATGCCAAGGCTTCTAGAGAAATGGGAAATGGGAAAACACCTGGGAA 1958
Qy 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1959 AACAAAGAAATCTTAAGAGAGACAGGGAAGAGTAAGGAAGTCTATAGTTGGAGATG 2018
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 2019 GAGAATGAGAATCTTGTAGAAAATGTTGCAGATCCGATCAAGATGATCAACAGCTTCTC 2078
Qy 695 LysGlnGlnSerProGlnProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2079 AACAAACAATCTCCAAAGAACCCAGTCTCTGAATGGTGGAGTTTGTAGAACACAC 2138
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2139 TTTCGTGAAGAAATCTACTACTCAGATCAGAAATCCAGAGATGTGAATCTGGAGGA 2198
Qy 735 GluValValLysGluLeuSerValGluGluGlnLeuLysArgAsnArgTrpTrpAspGlu 754
Db 2199 GAAGTGGTCAAGAGCTCTGTGGAAGAACACAGATAAGAGAAATCGGTATTATGATGAG 2258
Qy 755 AspGluAspGluGlu 759
Db 2259 GATGAGGATGAAGAG 2273
RESULT 6
BC001247
LOCUS
DEFINITION
Homo sapiens, epithelial protein lost in neoplasm beta, clone
MGC:4969 IMAGE:3452714, mRNA, complete cds.
ACCESSION
BC001247
VERSION
BC001247.1 GI:12654808
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2164)
Strausberg,R.
Direct Submission
Submitted (11-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: JRAK Plate: 4 Row: 0 Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7020350.
Location/Qualifiers
1. 2164

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:51474"
/db_xref="taxon:9606"
/clones="MGC:4969 IMAGE:3452714"
/tissue type="placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="vector: pCMV-Sport6"
187..1983
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAH01247.1"
/db_xref="GI:12654809"
/translation="MENCLEBSRHEVEKSEISENTDASGKIEKYVFLNRLKMMPEKG
EPTQTKILRAQSRGAKISENSYLDLEIGQLSSSTFDSKNEKSRRLNELPRL
SETSIKORMAKYQAAVSKQSSNTYNTELKASGGEIKHKWOKENYVPPGVEICITHQ
EGEKISANENSLAVRSTPADSDSDQVSKSEVQVPHKPLSPDRASSLSESSPPKA
MKPQAPARETCVCEQKTVYMERLLANQVPHISCFRCVCYNKLSLGTYSALHGRYI
CFHFNQLFKSGNYDEGFGPHKDLWASKNENEEILERPQLANARETPHSPGVED
APIAKVGLVAKAMEKASQOEKEDPAETKLLRIAMPPTTELSSGSALEEGIKMSK
KPWPPEBISKEPEVDVLDLKKLRSSSLKRSRPTVAASQSTSVKSPKTVSP
IRKMSQSESESVGGVRAERKOVENAKAKNGVGTWONKSGEGERKSGKEG
HSLMENENLVENGADSDDEDNLFKQSQPCKSLNWSFVDNTFAEFTTQNKQSK
DVEUWEVAVKELSVVEQIKRNYDEDEEE"

CDS

BASE COUNT 744 a 472 c 513 g 435 t
ORIGIN
Alignment Scores:
Pred. No.: 1,49e-178 Length: 2164
Score: 3417.50 Matches: 660
Percent Similarity: 99.85% Conservative: 0
Best Local Similarity: 99.85% Mismatches: 0
Query Match: 87.03% Indels: 1
DB: 9 Gaps: 1
US-09-890-549-4 (1-759) x BC001247 (1-2164)

Qy 99 HisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAla 118
Db 1 CACAGAGCAGACCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Qy 119 AspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuVal 138
Db 61 GACCAAGAGACAAATCCACCAGATCTAGATCTAGATCTAGATCTAGATCTAGATCTAGATCT 120
Qy 139 GlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSer 158
Db 121 CAGGTCGATATCCACATCAAGACCGTGAGGATCTTAAAGACCATCTCAACAGAAAGT 180
Qy 159 LysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSer 178
Db 181 AAAAAAATGGAAAAATTTGTAGGAGAAATCCAGGATCAAGTAAAGTAAAGTAAAGTAAAGT 240
Qy 179 GluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLys 198
Db 241 GAAAAACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAACAGGCTTAAG 300
Qy 199 MetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSer 218
Db 301 ATGATGTTTGAGAAAGGTGAACCAACTCAACTCAACTCAACTCAACTCAACTCAACTCAACT 360

QY 219 AlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyPro 238
 DB 361 GCAAGTGAAGGAAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCCA 420
 QY 239 GlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgAspAsnLeuGlu 258
 DB 421 GGTCAAGTGTCTATCTTCTACATTTGACTCGGAGAAAATGAGAGTAGACGAAATCTGGAA 480
 QY 259 LeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaVal 278
 DB 481 CTTCCAGCCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAGGCGCTGTG 540
 QY 279 SerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluIle 298
 DB 541 TCCAAACAAACAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATC 600
 QY 299 LysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThr 318
 DB 601 AAAATTCTAATAATGGACAAAAGAGATGTGCCCCAGGTCTGAGGTCTGGATCACC 660
 QY 319 HisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrPro 338
 DB 661 CATCAGGAAGGGAAAAGATTCTTCGAAATGAGAATAGCTGGCAGTCCGTTCCACCCCT 720
 QY 339 AlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnProValHisPro 358
 DB 721 GCCGAAGATGACTCCCGTGACTCCAGGCTTAAGAGTGAAGTTCACAGCCGTGTCATCCC 780
 QY 359 LysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAla 378
 DB 781 AAGCCACTAAGTCCAGATCCAGAGCCCTCCAGTCTTTCTGAAAGTCTCTCCGCCAAGCA 840
 QY 379 MetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyr 398
 DB 841 ATG---AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGATGTCAGAGACAGTCTAT 897
 QY 399 ProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCysPheArgCysSer 418
 DB 898 CCAATGGAGCGCTCTTTGGCCAAACAGCAGGTGTTTCACATCAGCTGCTCCGTTGCTCC 957
 QY 419 TyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCys 438
 DB 958 TATTGCAACAAACAACCTCAGTCTAGGAACATATGCATCTTTACATGGAAGAACTATTGT 1017
 QY 439 LysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHis 458
 DB 1018 AAGCTCACTTCAATCACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGCAC 1077
 QY 459 ArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgPro 478
 DB 1078 AGACCACACAGGATCTATGGGCAAGCAAAAATGAAACGAAGAGATTTTGGAGAGACCA 1137
 QY 479 AlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIle 498
 DB 1138 GCCCAGGTGCAATGCAAGGGAGACCCCTCACAGCCCGAGGGTAGAAGATGCCCTATT 1197
 QY 499 AlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLys 518
 DB 1198 GCTAAGTGGGTGCTCGCTGCTGCAAGTATGGAACCAAGCCCTCTCTCAGCAGGAGAAG 1257
 QY 519 GluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeu 538
 DB 1258 GAACACAAGCCAGCTGAAACCAAGACCTGAGGATCGCTGGCCACCCCACTGAACCTT 1317
 QY 539 GlySerSerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpProPro 558
 DB 1318 GGAAGTTCCAGGAGTGGCTTGGAGAGGAGGATCAAAATGTCAAAGCCCAATGGCCTCT 1377
 QY 559 GluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLeu 578
 DB 1378 GAAGACGAAATCAGCAAGCCGAAAGTTCCTGAGGATGCTAGATCTGAAAGAGCTA 1437

QY 579 ArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGln 598
 DB 1438 AGACGATCTTCTTCTACTGAGGAAAGAGCCGCCCATTCCTGTAGCAGCTTCATTTCAA 1497
 QY 599 SerThrSerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMet 618
 DB 1498 AGCAGCTCTGTCAAGAGCCCAAAACTGTGTCCCACTATCAGGAAAGGCTGGAGCATG 1557
 QY 619 SerGluGlnSerGluLysSerValGlyGlyArgValAlaGluArgLysGlnValGluAsn 638
 DB 1558 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCAGAAAGGAAACAAGTGGAAAT 1617
 QY 639 AlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSer 658
 DB 1618 GCCAAGGCTTCTAAGAAGAAATGGAAATGTGGGAAAAACAACCTGGCAAAACAAGAATCT 1677
 QY 659 LysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsn 678
 DB 1678 AAAGGAGACAGCAGGAGAGAGTGAAGAGGTGATAGTTTGGAGATGGAGATGAGAAT 1737
 QY 679 LeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSer 698
 DB 1738 CITGTAGAAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTCAACAACAATCT 1797
 QY 699 ProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGlu 718
 DB 1798 CCACAAGAACCCCAAGTCTCTGAATTTGGTGGAGTTTGTAGACAACACCTTTCTGTAAGAA 1857
 QY 719 PheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValLys 738
 DB 1858 TTCACTACTCAGAAATCCAGAAATCCCAAGGATGGAACTCTGGAGGGAGAGTGGTCAA 1917
 QY 739 GluLeuSerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluAspGlu 758
 DB 1918 GAGCTCTCTGTGGAGACACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 1977
 QY 759 Glu 759
 DB 1978 GAG 1980

RESULT 7
 AK000335 3277 bp mRNA linear PRI 22-FEB-2000
 LOCUS Homo sapiens cDNA FLJ20328 fis, clone HEP10039.
 DEFINITION
 ACCESSION AK000335
 VERSION AK000335.1 GI:7020350
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
 Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
 Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 3277)
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 Direct Submission
 Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,
 Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
 Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing; Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES		Location/Qualifiers	
source		1..3277	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="HEP10039"	
		/cell_line="HepG2"	
		/cell_type="hepatoma"	
		/clone_lib="HEP"	
		/notes="Cloning vector pME18SFL3"	
CDS		179..1978	
		/notes="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAA91092.1"	
		/db_xref="GI:7020311"	
		/translation="MENCIGSRHEVEKSEISNTDASOKIEKYNVPLNRLQMFEEK EPTQKILRAQSRGKRSISNSYSLDLEIGPQLSSSTFDSEKNRSRNLPEPLR SETIKDRMAKTAQAAVSKSSSTNYTNELKSGGEIKHKMEQKNVPPGPEVCITH EGEIKSANESLAVGTRDPAEDSDSOVKSEVQOPVHPKPLSPDRASLSSESPKA MKKFOAPARETCVCCQKTYVPMERLLANQOVFHISCFRCSCYCNKLISLGTYSLAGRI YCKPHNQLFKSGNVDGFGHRPHKDLWASKNEEILERPAQJANARETHPSGVE DAPIAKGVLAASMEAKASQOEKDKPAETKXLRAMPPTTELGGSSALBEGIKMS KPKPPDEISKPEVEDVDLKLRRSSLSKRSRPTVAASQSTSVKSPKTVSP PIRKGMSEQSEESVGGVABRKQVENAKSKNGVGTWONKESKGETGRSKE GHSLENNELVENGADSDDDNSFLKQSPQEPKSLNWSFVDNFAEEFTTQKRS QDVELWEGEVVLSVEEQIKRNYDEDEDESE"	
BASE COUNT		1089 a 651 c 703 g 834 t	
ORIGIN		US-09-890-549-4 (1-759) x AK000335 (1-3277)	
Alignment Scores:		Pred. No.: 2 94e-178 Length: 3277	
		Score: 3416.00 Matches: 658	
		Percent Similarity: 100.00% Conservative: 0	
		Best Local Similarity: 100.00% Mismatches: 0	
		Query Match: 86.99% Indels: 0	
		DB: 9 Gaps: 0	
QY		102 AspHisProAlaGluValThrSerHisAlaAlaSerGlyValAlaLysAlaAspGlnGlu 121	
Db		2 GACCAATCTCTGCTGAAGTGACACCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGNA 61	
QY		122 GluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArg 141	
Db		62 GAACAAATCCACCCAGATCTAGACTCAGCTACCTCTCTGAAAGCCCTCGTTTCAAGGTCGA 121	
QY		142 TyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMet 161	
Db		122 TATCCCAATCAAGACCGCTGAGGATCTTAAAGACCCTCAACAGAAAGTAAAAAATG 181	
QY		162 GluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThr 181	
Db		182 GAAATTTGCTAGGAGATCCAGGATGAAGTAGAAATATCAGAAATCAGTGAACACACA 241	
QY		182 AspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPhe 201	
Db		242 GATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTCAACAGGCTTAAGATGATGTT 301	
QY		202 GluLysGlyValProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSerGly 221	
Db		302 GAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGGCCAAAGCCGAAGTCAAGTGA 361	
QY		222 ArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGlnLeu 241	
Db		362 AGAAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCCAAGGTGAGTTG 421	
QY		242 SerSerSerThrPheAspSerGluLysAsnGlnSerArgArgAsnLeuGluLeuProArg 261	
Db		422 TCATCTTCTACATTTGACTCGGAGAAATAGAGTAGACGAAATCTGGAACTTCCACGC 481	
QY		262 LeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLysGln 281	
Db		1562 AGTGAGAGATCTGTGGTGGAGAGATTGCAGAAAGGAAACAGTGGAAATGCCAAGGCT 1621	
Db		1561 GTCAAGAGCCCAAAACTGTGTCCCACTTATCAGGAAAGGCTGGAGCATGTCCAGAGCAG 1561	
QY		602 VallySerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGluGln 621	
Db		1501 TCTTCACTGAAGGAAAGAGCCGCCATTCACGTGTAGCAGCTTCATTTCAAGACCTCT 1501	
QY		582 SerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSer 601	
Db		1441 ATCAGCAAGCCCAAGTTCTCTGAGGATGTGATCTAGATCTGAAGAGAGCTTAAGACGATCT 1441	
QY		562 IleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArgSer 581	
Db		1381 GGAAGTCCCTTGGAGGAGGAGATCAAAATGTCAAGCCCAATGGCTCTCTGGAAGAGCAA 1381	
QY		542 GlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProGluAspGlu 561	
Db		1321 CCAGTGAACCAAGAGCTGAGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCA 1321	
QY		522 ProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGlySerSer 541	
Db		1261 GGTGTCTGCTGCTCAAGTATGTAAGCCCAAGCCCTCTCTCAGCAGGAGAGAGAGCAAG 1261	
QY		502 GlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAspLys 521	
Db		1201 GCAATGCAGGGAGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTTATTTGCTAAGGTG 1201	
QY		482 AlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLysVal 501	
Db		1141 AAGATCTATGGGCAAGCAAAATGAAACGAAAGAGATTTTGGAGAGACCCAGCCAGCTT 1141	
QY		462 LysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGlnLeu 481	
Db		1081 CGTCTCTTGGCCCAACCAGCAGGTGTTTACATCAGCTGCTTCCGTGCTCTTATTGCAAC 961	
QY		442 AsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHis 441	
Db		1021 AACAACTCAGCTTAGGAACATATGCATCTTTTACATGGAAGAACTTATTGTAAGCCTCAC 1021	
QY		442 PheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHis 461	
Db		1081 TTCAATCAACTCTTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACAC 1081	
QY		402 ArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCysAsn 421	
Db		902 CGTCTCTTGGCCCAACCAGCAGGTGTTTACATCAGCTGCTTCCGTGCTCTTATTGCAAC 961	
QY		422 AspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeu 361	
Db		781 GACTCCCGTACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTA 781	
QY		362 SerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLys 381	
Db		841 AGTCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAG 841	
QY		382 PheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMetGlu 401	
Db		842 TTTTCAAGCACCCTGCAAGAGAGACCTGCGTGGAAATGTGAGAGAGAGCTTATCCCAATGGAG 901	

QY 642 SerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGlu 661
Db 1622 TCTAAGAGAGATGGGAATGTGGGAAACCAACCTGGCAAAACAAAGAACTTAAGAGAG 1681
QY 662 ThrGlyLysArgSerLysGlyGlyHisSerLeuGluMetGluAsnGluAsnLeuValGlu 681
Db 1682 ACAGGAGAGAGATGAAGAGTCAAGTTTGGAGATGGAGATGAGAACTTTGGAA 1741
QY 682 AsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGlnGlu 701
Db 1742 AATGGTSCAGACTCCGATGAAGATGAATACAGCTTCTCAACACCAATCTCCACAGAA 1801
QY 702 ProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheLaGluGluPheThrThr 721
Db 1802 CCCAAGTCTCTGAATGGTGCAGTTTGTAGACAAACCTTTGCTGAAGAATCACTACT 1861
QY 722 GlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeuSer 741
Db 1862 CAGAAATCAGAAATCCAGGATGTGGAATCTGGAGGGGAGAGTGGTCAAGAGCTCTCT 1921
QY 742 ValGluGluGlnLysLysArgAsnArgTyArgAspGluAspGluGluGlu 759
Db 1922 GTGGAAGACAGATTAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 1975

RESULT 8
BC031490 3232 bp mRNA linear ROD 20-SEP-2002
LOCUS Mus musculus, epithelial protein lost in neoplasm, clone MGC:27894
DEFINITION IMAGE:3498139, mRNA, complete cds.
ACCESSION BC031490
VERSION BC031490.1 GI:21619380
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3232)
Strasbourg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: capbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunsaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 35 Row: 1 Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12746445.

Location/Qualifiers
1. 3232
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:27894 IMAGE:3498138"
/tissue_types="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."

/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
81. 2342
/codon_start=1
/product="epithelial protein lost in neoplasm"
/protein_id="AAH31490.1"
/db_xref="GI:21619381"
/db_xref="LocusID:65970"
/translation="WESTPFRNRROWTSLSRVTAPELKVNNKSSAIVEIFSKYOKA
AEEANMERKKNPESLPQHFRRGTLVLKXKWPVAGAEPTDLSLNSSEGGHTAD
YPPAEVTPKPGVADREHTQKPRFGRSPEAVIQSRYPRSENHDPKQATSEKQ
MENCUGDRHEAEKPESTETSGIKYVPLNRLKMMFEKGEHNTQKSLWTSQRN
AGGRLLSNCSLDWEIGAGLSNFAKSEKSNLELPLRISETIKORMARYQA
AVSGDTPASYNELKTSKTHKWEQENVPGPPEACSVHQEGSKVSTTNSLVALS
VPAEDTSCNSVSKSAQOPHPKPLSPDARTSSLPSSPSKTAKKFOAPAKESCVECO
KTYVMEELLANQOVFHSICRSCYCNKLSLGTYSLSHGRLYCKPFRNQFLPKSGNY
DEGFGHKQKOLWASKSDNEETLGRPAQPPNAGESPHSPGVEDAPIAKGVLAASMEA
KASQREDEDPAETKLRIAWPPPAELGGSGSALLEGIKVSKPKWPPEDDVCKTEAP
EDVDLKLKRRSSLSKSRPFTVAASFTSSIKSPKASPSLRKMSSEGESEEFQ
GGIATMERKOTENARPSCGEKENVGSRWQGEVPRSKDRSSFELESENFPMGANIAE
DDNHVHAQOSPLEPEAPQWGSFVDTTAAKEFTTQNKRSQDVGFWEGEVRELSVEEQI
KRNRYDEDEEE"
BASE COUNT 969 a 792 c 853 g 618 t
ORIGIN

Alignment Scores:
Pred. No.: 7,936-154 Length: 3232
Score: 2970.00 Matches: 577
Percent Similarity: 83.82% Conservative: 60
Best Local Similarity: 75.92% Mismatches: 115
Query Match: 75.63% Indels: 8
DB: 10 Gaps: 3

US-09-890-549-4 (1-759) x BC031490 (1-3232)
QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 81 ATGGAATCGACTCCATTTAATAGACCCAGTGGACTTCCTGTCTTACGAGTAAACGCC 140
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 141 AAAGAGCTTCTCTTGTCAACAGAAACAAGTCATCCGCAATTTGTGAAATATTCTCCAAG 200

QY 41 TyrGlnLysAlaAlaGluGluThrAsnValLysLysArgSerAsnThrGluAsnLeu 60
Db 201 TACCAGAGGCTGCTGAAGAACCCCAACATGGAAGAGAAAGAAATAACCCCTGAAAGTCTG 260
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 261 CCCAGCAGCTTTAGAGAGAGGAGCCCTGCTGTGTTAAAGAGAGAGTGGGAGAACCCCGTG 320

QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 321 GCTGGGGCAGAAATCCACAGACTCTACTGCCAACACAGCAGCAGTGGGGTGGGCACACA 380
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 381 GCGGACTACCTCTGCTGAAGTACGACCAAGCCTGCTCTCTGGAGTCAAGTCAAGTCAAGCGG 440

QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 441 GAAGAGCACACCCAGCCCAACCTAGATTTCGATCTCTCCGGAAGCAGTATCCAGAGC 500
QY 141 ArgTyrProHisIleLysAspGlyGluAspLysAspHisSerThrGluSerLysLys 160
Db 501 CGGTATCTCTCGCTCAGAGAACAGCCACGATTTTAAAGCCAGCCAGCCAGGAGCAAAA 560
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 561 ATGGAATACTGTCTGGAGATTCAGGAGCATCAACAGAGAGAGCCAGAGAGCGAGGANAAC 620
QY 181 ThrAspAlaSerGlyLysIleGluLysTyAsnValProLeuAsnArgLeuLysMetMet 200

Db 621 ACAGAACTTCAGGCAAAATAGAGAAATACAGTTCCTCCTAGTATAGATGAGATGATG 680
QY 101 PheGluLysGlyGluProThrGlnThrLysIleLeuAArgAlaGlnSerArgSerAlaSer 220
Db 681 TTTGAGAAAGGTGAACACAAACCAACCAAGAGTCTCTGGAGCCCAAGCCCAAAATGCGGT 740
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlnGln 240
Db 741 GGAAGAGAGGTCTCTGAAACCAACTGTTCCCTGGATGACTGGGAAATAGTGGCGGTCA 800
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 801 TTGCTCTCTCTCCTCACTCACTCGGAGAAACAGAGATAGAGGATCTGAGAGTGCAC 860
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 861 CGCCTCTCGAAACCTCCATAAAGGACCGCATGGCCAAAGTACCAGGCTGCAGTGTCCAAG 920
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluLysIle 300
Db 921 CAGAGCAGCCAGCCAGCTACCAAAATGAGCTGAAACCAAGT-----GAAAGCAAAACT 974
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 975 CATAAATGGACAGAGAGAGATGTGCCCCAGGTCCCGAGGCTCGAGGCTGCATCAG 1034
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1035 GAAGGAAGCAAGGTTCTCAACTGAGATAGTCTAGTGGCCCTCTCTGTCCTGCTGAA 1094
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnProValHisProLysPro 360
Db 1095 GATGACACCTGTAACCTCCAGAGTGAAGAGCGAGGCCAGCGCTATGCACCCCTAAGCCG 1154
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1155 CTGAGTCTGAGCCAGGACCTCCAGTCTCCGGAAGTCTCCTTCCAAAACAGCGAAG 1214
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1215 AAGTTTCAGCGCCGCGCAAGAGAGAGCTCGTGGAGTGTCAAGAACCGGTGTACCCCATG 1274
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisLeSerCysPheArgCysSerTyrCys 420
Db 1275 GAACGCTCTCTGCGCAACAGCAGAGTGTTCACATCAGCTGTTCCGATGCTCTCTACTGC 1334
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1335 AACAAACAGCTCAGTCTAGGACATATGATCCTTACATGGACGATCTACTGTAAAGCT 1394
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1395 CACTTCAATCACTCTTCAAGTCCAAAGGCACTATGACGAGGCTTTGGGCAACAGCAG 1454
QY 461 HisLysAspLeuTyrAlaSerLysAsnGluGluIleLeuGluArgProAlaGln 480
Db 1455 CATAAGGATCTGTGGCAAGCAAGACACATGAGGAGACTTTGGGAGACCCAGCCAG 1514
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1515 CGCCTAATGCGGGAGAGCCCGCCATAGCCGGGTGTAGAAGATGCCCCCATGCCAAG 1574
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
Db 1575 GTCGCGTGTGGCGGCAAGTATGGAACCCAAAGGCTCTTCTCAGAGGGAGAGGAGAT 1634
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySer 540
Db 1635 AAGCTCGGGAGACCAAGAGCTGAGGATCGCTGGCTCCGCCAGCGAGCTGGCGGT 1694
QY 541 SerGlySerAlaLeuGluGlyLysMetSerLysProLysTyrProProGluAsp 560

Db 1695 TCCGGAAGTGCCTGGAGGAAGGATCAAGATTCGAAGCCCAAGTGCCTCCGGAGGAT 1754
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db 1755 GAGCTCTCAAGACGAGGAGCCCGCGAGGATGTAGATCTGGACCTGAACAACCTCGCGCG 1814
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1815 TCTTCGTCCTCAAGAGAGAGGAGCGCCGCTTACGCTAGCAGCTTCCTTTGAACTCT 1874
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTyrSerMetSerGlu 620
Db 1875 TCCATCAAGAGCCCAAGGCTCTTGGCCATCTCTCAGAAAGGTGGAGCGAGTCCGAG 1934
QY 621 GlnSerGluGluSerValGlyGly---ArgValAlaGluArgLysGlnValGluAsnAla 639
Db 1935 CAGAGTGAAGGATTTGGAGGAGGATAGCAGCAGATGGAAGGAAACAAACCGAAATGCC 1994
QY 640 LysLalaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
Db 1995 AGACCTCTGGGAGAAAGAGAATGTGGAAATACACGCTGGCAG----- 2039
QY 660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
Db 2040 GCGGAGGAAGTCCCGAGGAGCAAGGACCGTAGTTCCTTTGAGCTGGAGAGTGAATTTT 2099
QY 680 ValGluAsnGlyAlaAspSerAspAspAsnSerPheLeuLysGlnSerPro 699
Db 2100 ATGGAATATGGAGCAACATAGCTGAAGATGACAAACCATGTCCACGACAGCTCCCG 2159
QY 700 GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719
Db 2160 TTAGAACCCGAGGCTCGGGTGTGCTGGCTTTGTACACACCGCGCTGTAAAGAAATTC 2219
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGlu 739
Db 2220 ACTACCCAGAAATCAAAATCCCAAGATGTCGGGTCTTGGGAGGAGAGTGGTCCGAGAG 2279
QY 740 LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2280 CTGTCTGTGGAGGAACAGATAAAGAGAAACCGGTACTATGATGAGGACGAGGATGAAGAA 2339
RESULT 9
AF307845 2370 bp mRNA linear ROD 02-MAR-2001
LOCUS Mus musculus epithelial protein lost in neoplasm-b (Eplin) mRNA,
DEFINITION complete cds.
ACCESSION AF307845
VERSION AF307845.1 GI:11127934
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 2370)
TITLE Maul.R.S., Sachi Gerbin, C. and Chang, D.D.
Characterization of mouse epithelial protein lost in neoplasm
(EPLIN) and comparison of mammalian and zebrafish EPLIN
JOURNAL Gene 262 (1-2), 155-160 (2001)
MEDLINE 21100452
PUBMED 11179679
REFERENCE 2 (bases 1 to 2370)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2000) Medicine, UCLA, 10833 Le Conte Avenue, Los
Angeles, CA 90095, USA
FEATURES
source Location/Qualifiers
1..2370
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1..2370
/gene="Eplin"

CDS 109..2370 /gene="Eplin" /note="cytoskeleton-associated LIM domain protein; EPLIN-b" /product="epithelial protein lost in neoplasm-b" /codon_start=1 /db_xref="GI:11127935" /translation="MESTPNNRQWTSLSRVTAKELSLVNKNKSAIVEIFSKYQKA ABEANMERKXNPNESLPQHFRRGTLSVKKXNENPVAGAEFHTDLSFNSSSEGHTAD YPAEVDTPKAPQVRADREHTQPKRFGSRPEAVIQSRYPRENSHDFKQAQTESOK MENCLGRSHEAKPEKTESNTSEKIEKYNVPLNKKMFEXGHEHNOTKSLWTQNRN AGRRLLSNCSLDDWEIGAGHLSGSAFNSKESKRNLELPLRSTSIKDRMAKQA AVSKQSPASYTHLTKSEKTHKWKQXENVPPEACSVHQEGSKVSTTENSVALS VPVEDTCNSQVSEAQQPWHPPKPLSDARTSSLPSSPSKTAKKQAFAKESCVBCQ KTVPMERLIANQQVSEFCRSCYNNKLSLTGYASLHGRIYCKHFNLQPKSKNY DEFGHRQKHLWASKSDNETLGRPAQPNAGESPSPGVEDAPAKGVLLAASMEA KASQREDEKPAETKLRITAMPAPAEALGGSGALPEGIKVPKPPEDDDCKTAP EDVLDLKLRLRRSSLIKERSRPTVAASFSTSIKPKKASSPLRKGWSESESEFEG GGTATMRKQKTENARSPGKXENGVKRWQGEVFRSKDRSSFELESENFMENGANIAE DDNRVHAQSPQLPEAPGWSGFGVDTTAAKEFTTQNKQSDVGFMEGEVRELSEVEQI KRRRYDEDEDEE"

BASE COUNT 689 a 589 c 682 g 410 t

ORIGIN

Alignment Scores: Pred. No.: 8.03e-154 Length: 2370 Score: 2967.00 Matches: 576 Percent Similarity: 83.82% Conservative: 61 Best Local Similarity: 75.79% Mismatches: 115 Query Match: 75.55% Indels: 8 DB: 10 Gaps: 3

US-09-890-549-4 (1-759) x AF307845 (1-2370)

QY 1 MetGluSerSerProPheAsnArgAtGlnTrpThrSerLeuSerLeuArgValThrAla 20

Db 109 ATGGAATCGACTCCATTATATAGACGCGCAGTGACATCTCCCTGTCATGTAGAGTAACAGCC 168

QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerLys 40

Db 169 AAGAGCTTTCTCTGTGTACACAGAACAAAGTCATCCGCAATTTGGMAATATTCTTCAAG 228

QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60

Db 229 TACCAGAAGCTGCTGAAGAAGCAACATCGAAGGAAGAAATAAACCTCGAAGTCTG 288

QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80

Db 289 CCCAGACATTTAAGAGAGGACCTGCTGTGTGTTAAAGAAAGTGGGAGAACCCCGGTG 348

QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100

Db 349 GCTGGGCGCAATTCACACAGACTCTACTGCCAAACAGACGAGTGGGTGGGCACACA 408

QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120

Db 409 GCGGACTACCTCTCTGTGAAGTGACGCAAGCCCTGCTCTCGAGTCAGAGTGCACCG 468

QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140

Db 469 GAAGAGCACACCCAGCCCAACCTAGATTGTTGATCTCGTCCGGAAGCAGTTATCCAGAGC 528

QY 141 ArgTyrProHisLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160

Db 529 CGGTATCTCTGCTCAGAGAACACCCAGATTATTAAGCCCGCCAGCGAGAGCCCAAAA 588

QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180

Db 589 ATGGAATACTGTCTGGAGATTCAGGATTCAGGATGAGCAGAGAGAGCAGACGACGGAAC 648

QY 181 ThrAspAlaSerGlyLysLysGlyLysTyrAsnValProLeuAsnArgLeuLysMetMet 200

Db 649 ACAGAAACTTTCAGGCAAAATAGAGAAATACAACTGTTCCACTGAATAGACTGAAGATGATG 708

QY 201 PheGluLysGlyGluProThrGlnThrLysLysIleLeuArgAlaGlnSerArgSerAlaSer 220

Db 709 TTTGAGAAAGGTGAACACAAACCAACAGAGTCTCTGGACCCAAACCGAAATCGGGT 768

QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGln 240

Db 769 GGAAGGAGGCTCTCTGAAACAACTGTTCCCTGGATGACTGGAAATAGGTGCGGTCAT 828

QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260

Db 829 TTGTCTTCTCTGCAATTCACTCGGAGAAACAGAGATAAGAGGAATCTGGAGCTGCCA 888

QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280

Db 889 CGCTCTCAGAAACCTCCATAAAGACCGCATGGCAGTAGTACCAGCTGCAGTGTCCAA 948

QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyLysIleLys 300

Db 949 CAGAGCAGCCAGCCAGCTACAAATGAGCTGAAACCACTGTTT-----GAAAGCAAACT 1002

QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320

Db 1003 CATTAATGGGAACAGAGAGAAATGTGCCCGAGTCCCGAGCTCGAGCTGCATCAG 1062

QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340

Db 1063 GAAGGAACAGAGTTCTTACAACTGAGAAATAGTCTAGTGGCCCTCTCTGTCCCTGCTGAA 1122

QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360

Db 1123 GATGACACCTGTAACTCCAGGTGAAGAGCGAGCCCGAGAGCTATGCACCTTAAGCG 1182

QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380

Db 1183 CTGAGTCTCTGAGCCAGGACCTCCAGTCTTCGGAAGTCTCTCTTCCAAACAGCGAAG 1242

QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400

Db 1243 AAGTTTCAGGCGCGGCAAGAGGAGCTGCGTGGAGTGTCAAGACGCTGTACCCCATG 1302

QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420

Db 1303 GAACCGCTCTGCGCCACACAGAGTGTTCATCATCAGCTGTTTCCGATGCTCTTACTGC 1362

QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440

Db 1363 AACAAAGCTCAGTCTAGGAACATATCATCTCTTACATGGACGAATCTACTGTAAAGCT 1422

QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460

Db 1423 CACTTCAATCAACTCTTCAAGTCCAAAGGCAACTATGACGAGGCTTTTGGGCAACAGCAG 1482

QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGln 480

Db 1483 CATTAAGATCTGTGGGCAAGCAAGAGCGCAATCAGGAGACTTTTGGGAGACCAAGCCAG 1542

QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500

Db 1543 CGCGCTAATCAGGGGAGAGCCCCATAGCCGGGTGTAGAAAGATGCCCATCGCCNAG 1602

QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGlyGluAsp 520

Db 1603 GTGGCGTGTGGGCGCAAGTATGGAAGCCCAAGGCTCTTCTCAGAGGAGGAGGAGGAT 1662

QY 521 LysProAlaGluThrLysLysLeuArgIleAlaIleProProProThrGluLeuGlySer 540

Db 1663 AAGCCCGGAGAGCAAGAGCTGAGATGCGCTGGCTCCGCCAGCGAGCTGGCGGT 1722

QY 541 SerGlySerAlaLeuGluGlyLysLysMetSerLysProLysProLysProProGluAsp 560

Db 1723 TCCGGAGTCCCTGGAGGAGAGGATCAAGATTCGAAGCCCAAGTGGCTCCGAGGAT 1782


```

Db 715 AATCTGGAGCTGCCAGCGCTCTCAGAAACCTCCATAAAGACCGCATGCGCCAGTACCAG 774
QY 276 AlaIaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 775 GCTGCAGTGTCAAGCAGCAGCCAGCCAGCAGCTACACAAATGAGCTGAAACACAGT--- 831
QY 296 GlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProGlyProGluVal 315
Db 832 ---GAAAGCAAACCTCAATAATGGAGCAGAGGAGAAATGCTCCAGGTCCCGAGGCC 888
QY 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
Db 889 TGCACGCTCCATCAGGAGGAGCAGGTTCTGCACTGAGATAAGTCTAGTGGCCCTC 948
QY 336 SerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 949 TCTGTCCCTGCTGAAGATGACACCTGTAACCTCCAGGTGAAGAGCGAGGCCAGCAGCT 1008
QY 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1009 ATGCCACCTTAAGCCCTCAGTCTGACCCAGGACCTCCAGTCTCCGGAAAGTTCTCCT 1068
QY 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1069 TCCAAACAGCAGAAAGTTTCAGCGCGCGCAAGAGAGAGCTGCTGAGTGTCAAGAG 1128
QY 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPhe 415
Db 1129 ACGGTGTACCCCATGGAACGCTCTCTGCGCAACCCAGCAGGTGTTTCACATCAGCTGTTT 1188
QY 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1189 CGATGCTCTCTAGCAACAAGCTCAGTCTAGGAACATATGCATCTCTTACATGAGCTGTTT 1248
QY 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1249 ATCTACTGTAGCTTCACTTCAATCACTCTTCAAGTCCAAAGGCACTATGACGAGCGC 1308
QY 456 PheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluLeuLeu 475
Db 1309 TTTGGGCACAGCAGCATAGGATCTGTGGCAGCAGCAGCAGCATGAGGACACTTTG 1368
QY 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
Db 1369 GGGAGACCCAGCCAGCCCTAATGAGGAGAGCTCCCATACCCCGGTGTAGAGAT 1428
QY 496 AlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1429 GCCCCCATCGCCAAAGCGCGCTGCTGCGCGCAAGTATGGAAGCCAAAGGCTCTTCTCAG 1488
QY 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProPro 535
Db 1489 AGGAGAGGAGGATAGCTTGGGAGNCCAGAGCTGAGATCGCTGCTGCTCCGCCA 1548
QY 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLys 555
Db 1549 CTTGAGTGGCAGTTCGGGAAGTGCCTGAGGAGGAGGATCAAGATCGAAGCCCAAG 1608
QY 556 TrpProProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1609 TGGCTCCGGAGGATGAGCTGCAAGCAGGAGCGCCCGGAGGATGTAGATCTGGACCTG 1668
QY 576 LysLysLeuArgArgSerSerLeuLysGluArgSerArgProPheThrValAlaAla 595
Db 1669 AAGAGCTGGCGCGCTTCTGCTCCCTGAAGAGAGGAGCGCCCGTTCCTGTAGCAGCT 1728
QY 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgLysGly 615
Db 1729 TCCATTGGAACCTCTTCCATCAGAGCCCGCAGACCTCTTCGCCATCTCTCAGGAAGGT 1788
QY 516 TrpSerMetSerGluGlnSerGluGluSerValGlyGly---ArgValAlaGluArgLys 634

```

```

Db 1789 TGGAGCAGTCCGAGCAGAGTGGAGGAGGATACCGAGCATGGAAAGAAA 1848
QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnValGlyLysThrThrTyrGln 654
Db 1849 CAAACGGMAATGCCAGACCTCTGCGGAGAAAGAGATGTGGAAATACACCTGGCAG 1908
QY 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1909 -----GGCAGGAAAGTCCGAGGAGCAAGACCGTAGTCTTTTGGAGCTG 1953
QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeu 694
Db 1954 GAGAGTGAGAATTTTATGGAAATGGAGCAACATAGCTGAAGATGACCAACCATGTCCAC 2013
QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTyrSerSerPheValAspAsnThr 714
Db 2014 GCACAGCAGTCTCCGTAGAACCCGAGGCTCCAGGTTCCTGCTTGTAGACACAC 2073
QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGly 734
Db 2074 GCTGCTAAAGAAATCTACTACCCAGAAATCAAAATCCAAAGATGTCGGGTCTCGGAGGGA 2133
QY 735 GluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2134 GAAGTGTCCGAGAGCTGTCTGTGGAGAACACAGATAAGAGAAACCGGTACTATGATGAG 2193
QY 755 AspGluAspGluGlu 759
Db 2194 GACGAGGATGAAGAA 2208

RESULT 11
AF218025 1754 bp mRNA linear PRI 01-OCT-2000
LOCUS Homo sapiens clone PF624 unknown mRNA.
DEFINITION AF218025
ACCESSION AF218025.1 GI:10441979
VERSION 1
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1754)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Yu,J. and Han,L.H.
Novel Human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1754)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Yu,J. and Han,L.H.
Direct Submission
Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-Tu Road, Shanghai
200032, People's Republic of China
FEATURES
Location/Qualifiers
1..1754
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PF624"
444..1349
/codon_start=1
/product="unknown"
/protein_id="AAG17267.1"
/db_xref="GI:10441980"
/translation="MKFQAPARETCVCQKTVYXMERLLANQQVFHISCLRKYCXN
KLSLGTASLHGRYCKPHFNOLFKSGNYDEGFGRPHKDLKWSKNEBILRRPAQ
LANARETSPGVEDAPITAKVGLVAASMEAKSQEKEKDPATKKLRIAPPTPEL
GSSSALREGIKMSKPKPPDEITSKPEVEDVLDLKKLRSSSLSKRSRPFVAAS
FQSTSVKPTVSPPIRKGSWSQSSEESVGRVAERKQVENAKSKNGNVGTTWQ
NKESGEGREVRKVIWVRMRIL"
CDS

```

BASE COUNT 596 a 377 c 426 g 361 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 3.4e-138 Length: 1754
Score: 2679.00 Matches: 523
Percent Similarity: 98.87% Conservative: 0
Best Local Similarity: 98.87% Mismatches: 5
Query Match: 68.22% Indels: 2
DB: 9 Gaps: 0

US-09-890-549-4 (1-759) x AF218025 (1-1754)

QY 232 AspAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsn 251
DB 2 GATGACCTGGAATAGCCCGGCTAGTTGTCATCTTCTACATTTGACCTGGAGAAAAT 61
QY 252 GluSerArgArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMet 271
DB 62 GAGAGTAGAGCAATCTGGAACTTCCAGCGCTCTCAGAACTCTATAAGGATCGAATG 121
QY 272 AlaLysThrGlnAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeu 291
DB 122 GCCAAGTACCAGGCGAGTGTGTCACCAACAAAGCAGCTCAACCAACTATACAAATGAGCTG 181
QY 292 LysAlaSerGlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProPro 311
DB 182 AAGCCAGTGGTGGCGAAATCAAAATTCATAAATGGACAAAGGAGAGATGTGCCCCCA 241
QY 312 GlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSer 331
DB 242 GGTCTCTGAGCTCTGCATCACCCTACAGGAGGAGGAGAAAGATTCTCCAAATGAGAAATG 301
QY 332 LeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGlu 351
DB 302 CTGGCAGTCCGTCTCCACCCTCGCGAAGATGACTCCCGTGACTCCCGAGCTTAAGAGTGG 361
QY 352 ValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSe 371
DB 362 GTTCAACAGCCTGTCCATCCCAAGCCCACTAAGTCCAGATCTCCAGAGCCTCCAGCTTTTC 421
QY 371 rGluSerSerProGlyAlaMetLysLysPheGlnAlaProAlaArgGluThrCysVa 391
DB 422 TGNAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGT 481
QY 391 lGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHi 411
DB 482 GGAATCTCAGACAGACTCTATCAATGAGCGTCTCTTGGCCACCCAGCAGGTGTTTCA 541
QY 411 sIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSe 431
DB 542 CATCAGCTGTCRCGTGTCCTATTCATGCAKCAACAACTCAGTCTAGGAACATATGCATC 601
QY 431 rLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAs 451
DB 602 TTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTAAATCTAAGGGCAA 661
QY 451 nTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAs 471
DB 662 CTATGATGAGGCTTTGGCCACAGCCACCAAGGATCTATGGGCAAGCAAAATGAAAA 721
QY 471 nGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerPr 491
DB 722 CGAAGAGATTTTGGAGAGACCCAGCCAGCTTGCATGCAATGCAAGGGAGACCCCTCAGGCC 781
QY 491 oGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaLaSerMetGluAlaLy 511
DB 782 AGGGGTAGAAGATGCCCTATTGCTAAGGTGGCTGCTCTGGCTGCAAGATGATGAGGCCAA 841
QY 511 sAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAl 531
DB 842 GGCCTCTCTCAGCAGGAGGAGGAGCAAGCCAGCTGAACCAAGAGAGCTGAGGATCGC 901

QY 531 aTrrProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMe 551
DB 902 CTGGCCACCCCCCTACTGAACCTTGGAAAGTTTCAGGAAGTCCCTTGGAGGAGGATCAAAAT 961
QY 551 tSerLysProLysTrrProProGluAspGluLysSerLysProGluValProGluAspVa 571
DB 962 GTCAAGAGCCCAATGCGCTCTCTGAAGAGCAATTCAGCAAGCCCGCAAGTTCTTGAGATGT 1021
QY 571 lAspLeuAspLeuLysLysLeuArgSerSerSerSerLeuLysGluArgSerArgProPh 591
DB 1022 CGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCAAGGAAGAAAGCCGCCATT 1081
QY 591 eThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPr 611
DB 1082 CACTCTAGCAGCTTCAATTCAAAGCACCTCTCTCAAGAGCCCAAACTGTGTCTCCACC 1141
QY 611 oIleArgLysGlyTrrPrrSerMetSerGluGlnSerGluSerValGlyArgValAl 631
DB 1142 TATCAGGAAGAGCTGGAGCATGTGAGACGAGTGAAGAGTCTGTGGGTGGAAGAGTTGC 1201
QY 631 aGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh 651
DB 1202 AGAAGGAAACAAGTGGAATGCAAGCTTCTAAGAGAAATGGAAATGTGGGAAATAC 1261
QY 651 rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlyHisSe 671
DB 1262 AACCTGGCAAAACAAAGATCTAAAGGAGACA -GGGAAGAGAGTAAGAGAGTCTATAG 1320
QY 671 rLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAs 691
DB 1321 TTTGAGATGGAGAAATGAGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAAGATGATA 1380
QY 691 nSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheVa 711
DB 1381 CAGCTTCTCTCAAAACAACAATCTCCACAAAGAACCCCAAGTCTCTGAATTTGTCAGTTTGT 1440
QY 711 lAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLe 731
DB 1441 AGACACACCTTTTGTGTGAGAAATTCATCTACTCAGAAATCAGAAATCCCAAGATGGAACT 1500
QY 731 uTrrGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTy 751
DB 1501 CTGGAGGAGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGTA 1560
QY 751 rTyrAspGluAspGluAspGluGlu 759
DB 1561 TTATGATGAGATGAGGATGAAGAG 1585
RESULT 12
BD160117
LOCUS BD160117 2207 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160117
VERSION BD160117.1 GI:27865875
KEYWORDS JP 2002191363-A/14960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2207)
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 14960 09-JUL-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/14960
PD 09-JUL-2002
PF 28-JUL-2000 JP 200280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUO IOTSUKI

PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (580)..(2050).
Location/Qualifiers
1..2207
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 692 a 438 c 512 g 565 t
ORIGIN
Alignment Scores:
Pred. No.: 2,37e-126 Length: 2207
Score: 2465.00 Matches: 471
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 62.77% Indels: 0
DB: 6 Gaps: 0
US-09-890-549-4 (1-759) x BD160117 (1-2207)
QY 289 AsnGluLeuLysAlaSerGlyGluIleLysIleHisLysMetGluGlnLysGluAsn 308
Db 638 AATGAGCTGAAGGCGAGTGGTGGCAATCAAAATTCATAAAATGGAGCAAGAGCAAT 697
QY 309 ValProGlyProGluValCysIleThrHisGlnGluGlyIleSerAlaAsn 328
Db 698 GTGCCCCAGGCTCTGAGGTCTGATCACCCTCAGGAGGGAAGATTTCTGCAAAAT 757
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 758 GAGATAGCTGGAGTCCTTCACCCCTGCCAGAGATGACTCCCGTGACTCCAGATT 817
QY 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
Db 818 AAGAGTGAGGTTCAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCTCC 877
QY 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 878 AGTCTTCTGAAGATTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 937
QY 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGln 408
Db 938 ACCTGGTGGNAATGACAGACAGTCTATCCAAATGGAGCGTCTCTGGCCAAACCCAGCAG 997
QY 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428
Db 998 GTGTTTCAATCAGCTGCTTCCTGCTCTCTATTCACCAACCAACTCAGTCTAGGAACA 1057
QY 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 1058 TATGCATCTTTACATGGAAGAAATCTATTAAGGCTCCTCAATCAACTCTTTAAATCT 1117
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 1118 AAGGGCAACTATGATGAAGGCTTTGGCAGAGACCCACAGAGGATCTATGGCAAGCAAA 1177
QY 469 AsnGluAsnGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1178 AATGAAACCAAGAGATTTTGGAGAGACCCAGCCAGCTTCCAAATGCAAGGGAGACCCCT 1237
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaSerMet 508
Db 1238 CACAGCCCGGGGTAGAAGATGCCCCCTATTGCTAAGGTGGTCTCTGGCTGCAAGTATG 1297
QY 509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 1298 GAAGCCAAAGGCTCTCTCAGCAGGAGAGGAGCAAGCCAGCTGAAACCAAGAGCTG 1357

QY 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1358 AGGATCGCTGGCCACCCCTCAACTTGAAGTTGGAAGTGCCTTGGAGGAAGG 1417
QY 549 IleLysMetSerLysProLysTrpProProGluAspGluIleSerLysProGluValPro 568
Db 1418 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCT 1477
QY 569 GluAspValAspLeuAspLysLysLeuArgArgSerSerSerLeuLysGluArgSer 588
Db 1478 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACCATCTTCTTCTACTGAAGAAAGAAC 1537
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1538 CGCCCATTCACCTGTAGCAGCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAACTGTG 1597
QY 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1598 TCCCACTTATCAGGAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTCTGGGTGGA 1657
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1658 AGATTGCGAAGGAAACAAGTGGAAATGCCAAGCTTCTAAGAAAGAAATGGGAATGG 1717
QY 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1718 GGAATAACAACCTGCGCAAAACAAGATCTAAGAGAGACAGAGGAAGAAAGTAAGGAA 1777
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1778 GGTCTAGTTTGGAGATGGAGAAATGAGAATCTTTAGAAAATGGTGCAGACTCCGATGAA 1837
QY 689 AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1838 GATGATAACAGCTTCTCTCAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTGGTGC 1897
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 1898 AGTTTGTAGACACACCTTTCTCTGAAGAAATTCATCTACTCAGAAATCAGAAATCCCAGGAT 1957
QY 729 ValGluLeuTrpGluGlyGluValValLysLysGluLeuSerValGluGluGlnLysArg 748
Db 1958 GTGGAATCTCGGAGGAGAGAGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2017
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2018 AATCGGTATTATGATCAGGATGAGGATGAAGAG 2050
RESULT 13
AK023649
LOCUS
DEFINITION
Homo sapiens cDNA FLJ13587 fis, clone PLACE1009246, weakly similar
to POLLEN SPECIFIC PROTEIN SF3.
ACCESSION
AK023649
VERSION
AK023649.1 GI:10435634
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Iehida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
NEO human cDNA sequencing project
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2207)
Isogai, T. and Otsuki, T.
AUTHORS
Direct Submission
TITLE

JOURNAL Submitted (23-AUG-2000) Takao Ieogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@ri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES Location/Qualifiers

source 1..2207

organism="Homo sapiens"

molecule="mRNA"

db_xref="taxon:9606"

clone="PLACE1009246"

tissue_type="placenta"

clone_lib="PLACE1"

notes="cloning vector: pME18SFL3"

890..2053

notes="unnamed protein product"

codon_start=1

protein_id="BAI4625.1"

db_xref="GI:10435635"

/translation="MEQKENVPPGPEVCITHQGEKISANENSLAVRSTPABDDSRDS QVKEVQOPVHPKPLSPDSRASSLSSPPKMKKFOAPARETCVCEQKTVYMERLL ANQOVPHISCFRCVYNNKLSLCTYASLHCRIYCKPHNQLFKSKNYDEGGRPHK DLAKXNENEILERAQLANARETHPSGVEDAPLAKYGLVLAASWEAKSSQBEKED KPAETKLRITAWPPIELGSSGALBEGINMSKPKWPPDESKPEVPEVDLDLKL RRSLSKERSRPTVAASFOSTSVKSPKTPVPPPIRGWMSBQSEBSVGRVAERQV ENAKSKONGNVGKTQWNEKSGETGKSKGHELEMEINLVENGADSDDDNSFL KQSPQEPKSLNWSFVDNFTAFBEFTTQNKSDVLEWGEVVKLSVEQIKRNYR DEDEDEE"

BASE COUNT 692 a 438 c 512 g 565 t

ORIGIN

Alignment Scores:

Pred. No.: 2,37e-126 Length: 2207

Score: 2455.00 Matches: 471

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 62.77% Indels: 0

DB: 9 Gaps: 0

US-09-890-549-4 (1-759) x AK023649 (1-2207)

QY 289 AsnGluLeuLysAlaSerGlyGlyGluIleHisLysMetGluGlnLysGluAsn 308

Db 638 AATGAGTGAACCCAGTGTGGCGAAATCAAAATTCATAAATGAGCAAGAGGAGAT 697

QY 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328

Db 698 GTGCCCCCAGGTCTGAGGTCTGCATCACCCATCAGGAGGGGAAAGATTCTGCAAT 757

QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348

Db 758 GAGAAATAGCTGGCATGCTCGTCCACCCCTGCGAAGATGATCCGCTGACCTCCAGGTT 817

QY 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368

Db 818 AAGAGTGAGGTTCAACAGCTGTCCATCCAGCCACTAAGTCCAGATTCCAGAGCCCTCC 877

QY 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 878 AGTCTTCTGAAAGTTCTCTCCAAAGCAATGAAGAAGTTTCAGGCACCTCCAGAGAG 937

QY 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGln 408

Db 938 ACTGCGTGGATGTCAGAGACAGTCTATCCAAATGAGCGTCTCTGSCCAACCCAGCAG 997

QY 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnLysLeuSerLeuGlyThr 428

Db 998 GTGTTTCACATCAGTGTCTTCGTTGCTCTTATTGCAACAACTCAGTCTAGGAACA 1057

QY 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448

Db 1058 TATGCACTTTACATGGAAGAATCTATTGTAAAGCTTCACTTCAATCAACTCTTTAAATCT 1117

QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysLysAspLeuTrpAlaSerLys 468

Db 1118 AAGGCAACTATGATCAAGGCTTTGGGCACACACACCAAGGTTCTATGGGCAAGCAAA 1177

QY 469 AsnGluAsnGluGluLeuLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488

Db 1178 AATGAAACGAAGAGATTGAGAGAGAGCCAGCCAGCTTGCATAATGCAAGGAGAGCCCT 1237

QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508

Db 1238 CACAGCCAGGCTTGAAGATGCCCTTATTGTAAAGTGGGTGCTCTGGCTCAAGATG 1297

QY 509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528

Db 1298 GAAGCCAGGCTCTCTCAGCAGGAGAGGAGGAGCAAGCCAGCTGAAACCAAGAGGCTG 1357

QY 529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548

Db 1358 AGGATCGCTGGCCACCCCTTCACTTGAAGTTCAGGAAGTGCCTTGGAGGAGGG 1417

QY 549 IleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluValPro 568

Db 1418 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCGAGTCTCT 1477

QY 569 GluAspValAlaSerLeuLysLysLysLeuArgArgSerSerLeuLysGluArgSer 588

Db 1478 GAGGATGCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAGGAGG 1537

QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608

Db 1538 CGCCCATTTCACTGTAGCAGCTTCAATTTCAAAGACCTCTGTCAAGAGCCCAAAACTGTG 1597

QY 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628

Db 1598 TCCCCACCTATCAGGAAAGGCTGGAGCATGTGAGAGAGATGAGAGATCTGTGGGTGGA 1657

QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGluVal 648

Db 1658 AGAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1717

QY 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668

Db 1718 GGAAGAAACAACTGGCAAAACAAAGAAATCTAAAGGAGAGAGAGAGAGAGAGAGAGAG 1777

QY 669 GlyHisSerLeuGluMetGluAsnGluValGluAsnGlyAlaAspSerAspGlu 688

Db 1778 GGTCAATGTTGGAGATGAGAGATGAGATCTTTGAGAAATGGTGAGATCTCCGATGAA 1837

QY 689 AspAspAsnSerPheLeuLysGlnSerProGlnGluProLysSerLeuAsnTrpSer 708

Db 1838 GATGATAACAGCTTCTCTCAAAACAAATCTCCCAAGACCCCAAGTCTCTGAATGGTGG 1897

QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728

Db 1898 AGTTTGTAGACACACCTTTGCTGAAGAATTCATCTACTCAGAAATCAGAAATCCAGGAT 1957

QY 729 ValGluLeuTrpGluGlyGluValValValValValValValValValValValValVal 748

Db 1958 GTGAACTCTGGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAGAGAGAGAGAGAGAG 2017

QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759

Db 2018 AATCGGTATTATGATGAGGATGAGGATGAGAG 2050

RESULT 14

AK096172

LOCUS

DEFINITION Homo sapiens cDNA FLJ38853 fis, clone MESAN2010321, highly similar

to Homo sapiens epithelial protein lost in neoplasm beta (EPLIN)
mRNA.
ACCESSION AK096172
VERSION AK096172.1 GI:21755594
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2793)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. 2793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MESAN2010321"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN2"
/note="Cloning vector: pME18SFL3-primary culture, normal
mesangial cells"
BASE COUNT 826 a 577 c 637 g 753 t
ORIGIN
Alignment Scores:
Pred. No.: 5,89e-126 Length: 2793
Score: 2460.00 Matches: 470
Percent Similarity: 99.79% Conservative: 0
Best Local Similarity: 99.79% Mismatches: 1
Query Match: 62.64% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x AK096172 (1-2793)
Qy 289 AsnGluLeuLysAlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsn 308
Db 1230 AATGAGCTGAAGCCAGGTGGCGAATCAATTCATAAAATGGAGCAAGAGGAAT 1289
Qy 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328
Db 1290 GTGCCCCAGGTCTGAGGTGTCATCCATCAGGAAGGGGAAAGATTTCTGCAAT 1349
Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 1350 GAGAAATAGCTGGGAGCTCGGTCCACCCCTCCCAAGATGACTCCCGTGAATCCCAAGTT 1409
Qy 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
Db 1410 AAGAGTGAGGTTCACAGCTGTCCATCCCAAGCCATAGTCCAGATTCAGAGCCCTCC 1469
Qy 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 1470 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTCTCAAGAGAG 1529
Qy 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 1530 ACCTGGTGGATGTGTCAGAGACAGTCTATCAATAGAGCGTCTCTTGCCCAACAGCAG 1589
Qy 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThr 428
Db 1590 GTGTTTTCATCAGTCTGCTTCCGTTCTCTTATGTCATTCGCAACAACTCAGTCTAGGA 1649
Qy 429 TyrAlaSerLeuHisGlyArgIleTyCysValProHisPheAsnGlnLeuPheLysSer 448
Db 1650 TATGCATCTTTACATGGAAGATCTATTGTAAGCTCTCACTCAATCAACTCTTTAAATCT 1709
Qy 449 LysGlyAsnTyAspGluGlyPheGlyHisArgProHisLysAspLeuThrAlaSerLys 468
Db 1710 AAGGGCAACTATGATGAGGCTTTGGGCACACACACCAAGATCTATGGGCAAGCAA 1769
Qy 469 AsnGluAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1770 AATGAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCATAATGCAAGGGAGAGCCCT 1829
Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 1830 CACAGCCAGGGGTAGAAGATGCCCTTATGTGAAGTGGGTCTCTCTGCTGCTCAAGATG 1889
Qy 509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 1890 GAAGCCAGGCTCTCTCAGCAGAGAGAGAGAGCAAGCCAGCTGAAACCAAGAGCTG 1949
Qy 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1950 AGGATCGCTGGCCACCCCTCACTGAAGTTCAGGAAGTGCCTTGGAGGAGAGG 2009
Qy 549 IleLysMetSerLysProLysTrpProGluAspGluLysSerLysProGluValPro 568
Db 2010 ATCAAAATGTCAAGCCCAATATGGCTCTCTGAAGAGCAATCAGCAAGCCCAAGTTCTCT 2069
Qy 569 GluAspValAspLeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSer 588
Db 2070 GAGGATGTGATCTAGATCTGAAGAGCTAAGCAGCCTTCTTCACTGAAGAGAGAG 2129
Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 2130 CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTCTCAAGAGCCCAAAACTGTG 2189
Qy 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 2190 TCCCACTATCAGGAAAGGCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTCTGGGTGA 2249
Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 2250 AGAGTTGCAGAAAGGAAACAGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTG 2309
Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 2310 GGAATAAACCTCGCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGTGAAGAA 2369
Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 2370 GGTGATAGTTTGGAGATGAGAAATCTTTAGAAAATGGTGCAGACTCCCATGAA 2429
Qy 689 AspAspAsnSerPheLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 2430 GATGATAACAGCTTCTCAACCAACAATCTCCCAAGAACCCCAAGTCTCTGAAATGGTTCG 2489
Qy 709 SerPheValAspAsnThrPheAlaGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 2490 AGTTTGTAGACAAACACCTTTCTGAAAGATTCCTACTCTCAGAAATCAGAAATCCAGGAT 2549
Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGlnLysArg 748

Db 2550 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2609

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759

Db 2610 AATCGGTATTATGATGAGGATGAGATGAAGAG 2642

RESULT 15

AX017487

LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000

DEFINITION Sequence 31 from Patent WO9947655.

ACCESSION AX017487

VERSION AX017487.1 GI:10042284

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilarsky,C.

TITLE Human nucleic acid sequences from normal breast tissue

JOURNAL Patent: WO 9947655-A 31 23-SEP-1999;

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN (DE); PILARSKY CHRISTIAN (DE)

FEATURES

Location/Qualifiers

1..2783

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 901 a 526 c 609 g 747 t

ORIGIN

Alignment Scores:

Pred. No.: 7.55e-126 Length: 2783

Score: 2458.00 Matches: 470

Percent Similarity: 99.79% Conservative: 0

Best Local Similarity: 99.79% Mismatches: 1

Query Match: 62.59% Indels: 0

DB: 6 Gaps: 0

US-09-890-549-4 (1-759) x AX017487 (1-2783)

Qy 289 AsnGluLeuLysAlaSerGlyGlyGluLeuLysLeuLysMetGluGlnLysGluAsn 308

Db 40 AATGAGCTGAAGCCAGTGGTGGCGAATCAAAATTCATAAATGGAGCAAGAGAGAT 99

Qy 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328

Db 100 GTCCCCCAGCTCTGAGGTCTGCATCCCATCCCATCAGGAAGGGGAAAGATTTCTGCAAT 159

Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348

Db 160 GAGAAATAGCTGCGAGTCCGTTCCACCCCTGCCGAGATGACTCCGTCGACTCCAGGTT 219

Qy 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368

Db 220 AAGAGTGAGGTTCAACAGCGCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCC 279

Qy 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 280 AGTCTTTCTGAAAGTCTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTCGAAGAG 339

Qy 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408

Db 340 ACCTGCGTGGATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCCAGAG 399

Qy 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428

Db 400 GTGTTTCACATCAGCTGCTTCGTTGCTCTATTTGCAACCAACAACTCAGTCTAGAACACA 459

Qy 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448

Db 460 TATGCATCTTTACATGGAAGAATCTATTTCAGACCTCACTTCAATCAACTCTTTAAATCT 519

Qy 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468

Db 520 AAGGGCAACTATCATGATGAAGGCTTTGGGCACACACACCAAGCATCTATGGGCAAGCAA 579

Qy 469 AsnGluAsnGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488

Db 580 AATGAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAATGCAAGGGAGACCCCT 639

Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508

Db 640 CACAGCCAGGGGTAGAGATGCCCTATTGTGAAGGGGGTGTCTTGCTGCAAGTAGT 699

Qy 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528

Db 700 GAAGCCAGGCTCTCTCAGCAGAGAGGAAGCAAGCCAGCTGAAACCAAGAGCTG 759

Qy 529 ArgIleAlaTrpProProProProGluLeuGlySerSerGlySerAlaLeuGluGly 548

Db 760 AGGATCGCTGGCCACCCCTCACTGAACTTGGAGTTTCAAGAGTTCAGGAGGAGGG 819

Qy 549 IleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluValPro 568

Db 820 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCT 879

Qy 569 GluAspValAspLeuAspLysLysLeuArgArgSerSerSerSerLysLysGluArgSer 588

Db 880 GAGGATGCGATCTAGATCTGAAGAGCTTAAGCAGATCTTCTTCACTGAAGAAAGAGC 939

Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608

Db 940 CGCCCATTCACCTGTAGCAGCTTCAITTTCAAAGCACCTCTGTCAAGAGCCCAAACTGTG 999

Qy 609 SerProProIleArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGly 628

Db 1000 TCCCCACCTATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGA 1059

Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648

Db 1060 AGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAGAGCAATGGGAATGTG 1119

Qy 649 GlyLysThrThrTyrGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668

Db 1120 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGAGAGACAGGAGAGAGAGTAAGGAA 1179

Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688

Db 1180 GGTCAATAGTTTGGAGATGAGCAATCAGAAATCTTGTAGAAAATGGTGCAGACTCCGATGA 1239

Qy 689 AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708

Db 1240 GATGATAACAGCTTCTCTCAACAAACAATCTCCCAAGAACCCCAAGTCTCTGAATTGGTGC 1299

Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGluAsp 728

Db 1300 AGTTTGTAGACAACACCTTTGCTGAAGAATTTACTACTCAGAAATCAGAAATCCAGGAT 1359

Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748

Db 1360 GTGGAATCTCTGGAGGGAGAGTGGTCAAGAGAGCTCTCTGTGGNAGAACACAGATAAGAGA 1419

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759

Db 1420 AATCGTATTATGATGAGGATGAGGATGAAGAG 1452

Search completed: January 7, 2004, 00:38:49

Job time : 7272 secs

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3927	100.0	2905	22	AAI58169	Human polynucleoti
2	3927	100.0	3664	23	ABX71307	Human transcriptio
3	3927	100.0	3705	21	AAAS3826	Sequence encoding
4	3909.5	99.6	3711	21	AACT7143	Human OSFX ORF2698
5	3872.5	98.6	3650	22	AAFS5697	Human EPLIN (epith
6	3603.5	91.8	3543	22	AAFS5696	Human EPLIN (epith
7	3285.5	83.7	3465	23	AAST7246	DNA encoding novel
8	2679	68.2	1754	24	AB198873	Human cancer suppr
9	2537	64.6	1713	22	AAI59955	Human polynucleoti
10	2535	64.6	2749	24	ABK35261	Human cDNA encodin
11	2465	62.8	2207	22	AAH18125	Human cDNA sequenc
12	2455	62.5	2783	22	AAK33566	Human breast tumou
13	2449	62.4	1567	24	AAI18588	CDNA encoding huma
14	1586	40.4	2267	21	AAI59489	Human secreted pro
15	1411	35.9	2158	21	AAI93483	Human secreted pro
16	1254	31.9	732	21	AAI98824	Human pancreatic c
17	881	22.4	641	20	AAI52892	Human prostate tum
18	841.5	21.4	698	22	AAH72864	Human cervical can
19	828	21.1	565	24	ABV98930	Human pancreatic c
20	815	20.8	1014	23	AAI88277	DNA encoding novel
21	615.5	15.7	4269	24	ABQ61002	Human breast tumou
22	606	15.4	413	25	AAI56082	Chinese hamster si
23	596.5	15.2	3825	24	AAI32329	Human lung specifi
24	541	13.8	2547	22	AAI33284	DNA encoding human
25	538	13.7	2336	22	AAI33341	DNA encoding human
26	534.5	13.6	2379	24	AAI42854	Human DNA #13 core
27	534.5	13.6	2379	24	AAI27210	Human 41441 protei
28	506.5	12.9	577	23	ABV49283	Human prostate exp
29	442.5	11.3	710	25	AAI56103	Chinese hamster si
30	423	10.8	532	22	AAI12979	Human cDNA clone (
31	414	10.5	389	21	AAI20709	Human secreted pro
32	392	10.0	1638	24	ABI70363	Human bone remodel
33	390	9.9	1848	23	AAI77023	DNA encoding novel
34	388.5	9.9	1321	24	ABL39708	Human NS cDNA sequ
35	385.5	9.8	2226	23	AAI72745	DNA encoding novel
36	385	9.8	1298	21	AAI27128	Human inflammation
37	384	9.8	408	23	AAI72740	DNA encoding novel
38	383	9.8	1122	24	ABL39709	Human NS cDNA sequ
39	383	9.8	1205	24	ABL39710	Human NS cDNA sequ
40	383	9.8	1284	24	ABL39707	Human NS cDNA sequ
41	379.5	9.7	1159	24	ABL39712	Human NS cDNA sequ
42	379.5	9.7	1242	24	ABL39711	Human NS cDNA sequ
43	374.5	9.5	654	25	AAI56396	Chinese hamster si
44	372	9.5	847	23	AAI56534	DNA encoding novel
45	370.5	9.4	1503	22	AAI58876	Human immune/haema

ALIGNMENTS

RESULT 1

AAI58169

ID AAI58169 standard; cDNA; 2905 BP.

XX AAI58169;

AC AAI58169;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 372.

DE

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

XX

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 15:28:21 : Search time 497 Seconds (without alignments)

4122.482 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVBEQIKRNYDEDEDEE 759

Scoring table:

BLOSUM62	
Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2.1/USPTO.spool/US09890549/runat.06012004.094752.19815/app.query.fasta.1.903

-DB=N_Geneseq.19Jun03 -QMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=closm62 -TRANS=human40.cdi

-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=prco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09890549 @CGN 1.1.312 @runat.06012004.094752.19815 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

OS Homo sapiens.
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0682191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX P-PSDB; AAM39013.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 372; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2905 BP; 965 A; 617 C; 691 G; 632 T; 0 other;

Alignment Scores:
Pred. No.: 1.25e-272 Length: 2905
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-890-549-4 (1-759) x AA158169 (1-2905)

QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 173 ATGGAAATCATCTCCATTATAACAGCGCAATGACCTCACTATCATTTGAGGGTAACAGCC 232

QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerTys 40
Db 233 AAAGAATCTTCTCTGTCACCAAGAAACAAGTCATCGGCTATTGTGAAATATTCTCCAAG 292

QY 41 TyrGlnLysAlaAlaGluGlnThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 293 TACCAGAAACGACCTGAGAGAAACAACATCGGAGAAAGAGAGAAAGTAACACCGAAATCTC 352

QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80

Db 353 TCCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGTGGGAGAACCCAGGG 412
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 413 CTGGGAGCAGAGTCTCACACAGACTCTCTACCGAAGCAGCAGCAGCTGAGATTAGGCACAGA 472
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 473 GCAGACCATCTCTCTGCTGAAGTGACAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAA 532
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 533 GAAGAACAATCCACCCAGATCTAGACTCAGCTCACCTCTGAGAGCCCTCTTCCAGGT 592
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 593 CGATATCCCCACATCAAGACCGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 652
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 653 ATGGAAATTTGCTAGGAGATCCAGGATGAAGTAGAAATCAGAAATCAGTGAAGAAC 712
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 713 ACAGATGCTTCGGCAAAATAGAGAAATATAATGTTCCGCTGACAGAGCTTAAGATGATG 772
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 773 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 832
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 833 GAAAGAAAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCCCAGGTGAG 892
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 893 TTGTCTATCTTCTACATTTGACTCGAGAGAAATAGAGTAGACGAAATCTGGAATTTCCA 952
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 953 CGCCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAGGAGCTGTGTCCAAA 1012
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluLysIle 300
Db 1013 CAAGACAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCAAAATCAAAAT 1072
QY 301 HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
Db 1073 CATAAATGGAGCAAAAGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCACCATCAG 1132
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1133 GAAGGGGAAAAGATTTCTGCAATGAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1192
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1193 GATGACTCCCGTGACTCCAGGTTAAGAGTGAAGTTCACAGAGCTGTCCATCCCAAGCCA 1252
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1253 CTAAGTCAGATTCAGAGCCCTCCAGTCTTCTGAAAAGTTCTCTCCCAAGCAATGAAG 1312
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1313 AAGTTTCAGGCACCTCGAAGAGACCTCGGTGAATGTGAGAGAGAGAGTCTATCCCAATG 1372
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1373 GAGCGTCTCTTGGCCAAACCCAGCAGGTCTTTCACATCAGTGTCTCCGTTCCGTTCTATTGC 1432
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440

1433 AACAAACAACTCAGCTAGGACATATGCATCTTTACATGGAGAAATCTATTGTAAGCCT 1492
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1493 CACTTCAATCAACTCTTTAAATCTTAAAGGCAACTATGATGAAGGCTTTGGGCACAGACCA 1552
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluLysLeuGluArgProAlaGln 480
Db 1553 CACAAGGATCTATGGGCAAGCAAAATGAAACGAGAGATTTTGGAGAGACCAAGCCAG 1612
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1613 CTGTCAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTTATTGCTTAAG 1672
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 1673 GTGGGTGCTGCTGGCTGAAGTATGAAGCAAGGCTCTCTCAGCAGGAGAGAGAC 1732
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
Db 1733 AAGCCAGCTGAAACCAAGAACTGAGGATCGCTGCCACCCCTGAACTTGAAGT 1792
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpProProGluAsp 560
Db 1793 TCAGGAAGTCTGTGGAGAGAGGATCAAATGTCAAAGCCCAATGGCTTCCCTGAAGAC 1852
QY 561 GluLysSerLysProGluValProGluAspValAspLeuLysLysLeuArgArg 580
Db 1853 GAAATCAGCAAGCCGAAGTCTCTGAGGATGTCGATCTAGATCTGAAGAGCTTAAGACA 1912
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1913 TCTTCTTCACTGAAGGAAAGAGCGCCCATCTACTGTAGCAGCTTCAATTCAAAGCACC 1972
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSerGlu 620
Db 1973 TCTGTCAAGACCCCAAAACCTGTGTCCCACTATCAGGAAGGCTGGAGCATCTCAGAG 2032
QY 621 GlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 2033 CAGAGTGAAGAGTCTGTGGTGGAGAGGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAG 2092
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGly 660
Db 2093 GCTTCTAAGAAAGATGGGAATGTGGGAAACAAACCTGGCAAAACAAAGAACTTAAGGA 2152
QY 661 GluThrGlyLysArgSerLysGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 GAGACAGGGAAGAGAGTAAGGAAGGTATAGTTGGAGATGGAGAAATGAGAAATCTTGT 2212
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGTGTGAGACTCCGATGAAGATGATTAACAGCTTCTCTCAACCAACAACTCTCCACA 2272
QY 701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCCAAAGTCTCTGAATGTGTGAGTTTGTAGACAAACCTTGTCTGAAGAAATTCAC 2332
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
Db 2333 ACTCAGAATCAGAAATCCAGGATGTGGAATCTGGGAGGAGAGAGTGTCTCAAGAGCTC 2392
QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluGlu 759
Db 2393 TCTGTGAAGAACAGATAAGAGAGAAATCGGTATTATGATGAGGATGAAGAG 2449

RESULT 2
ABX71307
ID ABX71307 standard; cDNA; 3664 BP.
XX
AC ABX71307;
XX
DT 14-APR-2003 (first entry)

XX Human transcription factor cDNA from clone DKFZphut1_18c19.
DE Human; gene; gene therapy; vaccine; disease treatment: detection; ss.
KW Homo sapiens.
OS WO200112659-A2.
XX 22-FEB-2001.
PD 18-AUG-2000; 2000WO-1B01496.
XX 18-AUG-1999; 99US-0149499.
PF 28-SEP-1999; 99US-0156503.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
PA Wiemann S;
XX WPI; 2001-327840/34.
PI P-PSDB; ABUS2869.
DR XX
CC Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies .
PS Claim 1; Page 449-450; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence encodes a polypeptide
CC described in the disclosure of the invention.
XX SQ Sequence 3664 BP; 1210 A; 734 C; 809 G; 911 T; 0 other;
Alignment Scores:
Pred. No.: 1 68e-272 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-890-549-4 (1-759) x ABX71307 (1-3664)
QY 1 MetGluSerSerProPheAsnArgArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGATCATCTCCATTTAATAGCGGCNATGAGCCTCACTATCATTTAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTCTCTTGTCAACAAGAACAGTCATCGCTATTGTGGAATATATCTCCAG 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTGAAGAAACAAACATGAGAGAGAGAGAGTAACACCAATCTC 273
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCAGACTTTTAGAAGGGGACCTGACTGTGTGTAAAGAGAGAGTGGAGAACCCAGGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCAGCTAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120

394 GCAGACCATCTCTGCTGAAGTGACAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAA 453
 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
 454 GAAGAAACAAATCACCCACAGATCTAGACTCAGGTCACTCTCTGAAGCCCTCGTTCAGGT 513
 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
 514 CGATATCCCCACATCAAGGACGCTGAGGATCTTTAAAGACCATCTCAACAGAAAAGTAAAAA 573
 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 574 ATGGAATTTGCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAATAC 633
 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 634 ACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAACAGGCTTAAGATGATG 693
 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 694 TTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGGAAGTCAAGT 753
 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
 754 GGNAGGAAGATCTCTGAACACAGCTATTCTTAGATGACTGTGAATAGGCCACAGGTGAG 813
 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
 814 TTGTGATCTCTACATTTGACTCGGAGAAAAATGACAGTAGACGAAATCTGGAATCTCCA 873
 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
 874 CGCCTCTCAAAACCTCTATAAGAGTGAATGGCCAGTACGAGCAGCTGTGTCCAAA 933
 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyLysIleLysIle 300
 934 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCGGAAATCAAAAT 993
 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 994 CATAAATGGAGCAAAAGCAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCAACCATCAG 1053
 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 1054 GAAGGGGAAAGATTTCTGCAATGAGATAGCTGGCAGTCCGTTCCACCCCTCCCGAA 1113
 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
 1114 GATGACTCCCGTACTCCAGGTTAAGAGTGAGGTTCACACGCTGTCCATCCCAAGCCA 1173
 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
 1174 CTAAATCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAG 1233
 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
 1234 AAGTTTCAGGCACCTGCAAGACAGACCTCGTGGGAATGTCAAGACACAGTCTATCCAATG 1293
 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
 1294 GAGCGTCTCTCGCCCAACAGCAGGTGTTTCAATCAGCTGCTTCCGTTGCTCTCTATTGC 1353
 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
 1354 AACACAAACTCAGTCTAGAACATATGATCTTTTACATGAAGAAATCTATTGAAGCCT 1413
 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
 1414 CACTTCAATCAACTCTTTAAATCTAAGGCCAATATGATGAAGCTTTGGCCACAGACCA 1473
 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
 1474 CACAAGGATCTATGGGCAAGCAAAATGAAACGAAGAGATTTTGGAGAGACCCAGCCAG 1533

QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
 Db 1534 CTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGTGAAGATGCCCTATTGCTTAAG 1593
 QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
 Db 1594 GTGGGTGCTCTGGTGCAGATATGGAAGCCCAAGGCTCTCTCAGCAGGAGAGGAAGAC 1653
 QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
 Db 1654 AAGCAGCTGAAACCAAGAGAGCTGAGGATCGCTGGCCACCCCTCACTGAACCTTGAAGT 1713
 QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProProGluAsp 560
 Db 1714 TCAGGAAGTGCTTTGGAGGAAGGATCAAAATGTCAAGGCCAAATGGCTCTCTGAAGAC 1773
 QY 561 GluLysSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArg 580
 Db 1774 CAATTCAGCAAGCCGGAAGTCTCTGAGGATGTGATCTAGATCTGAAGAGCTTAAGACGA 1833
 QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
 Db 1834 TCTTCTCACTGAAGGAAGAGAGCCGCCATTCCTAGCAGCTTCAATTCGAAGCACC 1893
 QY 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
 Db 1894 TCTGTCAAGACCCCAAAATCTGTCTCCACCTATCAGGAAGGCTGGAGCATGTCAAG 1953
 QY 621 GlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
 Db 1954 CAGAGTCAAGAGTCTGTGGTGGAGAGAGTTCGCAAGAGGAAACAAGTGGAAAATGCCAAG 2013
 QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
 Db 2014 GCTTCTAAGAAGAAATGGAAATGTGGGAAAAACAACCTGGCAAAACAAGAATCTTAAAGGA 2073
 QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
 Db 2074 GAGACGGGAGAGAGAAATGAAGAGAGTCAAGTTCGAGATGGAGAAATGAGAATCTTSTA 2133
 QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
 Db 2134 GAAATGTCGACACTCCGATGAAGATGATPAACAGCTTCTCTCAACAACTATCCACAA 2193
 QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
 Db 2194 GAACCAAGTCTCTGAATTCGTGAGTTTGTAGACACACCTTTGCTGAAGAAATTCCT 2253
 QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
 Db 2254 ACTCAGAATCAGAAATCCAGGATGTGAACTCTGGAGGGAGAGTGTCAAAGAGCTC 2313
 QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
 Db 2314 TCTGTGGAGAACAGATTAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2370
 RESULT 3
 ID AAA53826
 XX AAA53826 standard; DNA; 3705 BP.
 AC AAA53826;
 XX
 DT 03-JAN-2001 (first entry)
 XX
 DE Sequence encoding lipid associated protein (LIPAP) 2766980CB1.
 XX Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
 KW antagonist; antibody; cardiovascular disease; neurological disease;
 KW gastrointestinal disease; lipid metabolism; detection;
 KW amplification; monitoring; hybridisation; antisense; triplex;
 KW ribozyme; screening; immunoassay; ds.
 XX

OS XX Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 137..2416
 FT /*tag= a
 FT /product= Lipid associated protein
 XX
 PN WO200049043-A2.
 XX
 PD 24-AUG-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04160.
 XX
 PR 19-FEB-1999; 99US-0120703.
 PR 08-JUL-1999; 99US-0142762.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Tang YT, Hallman JL, Yue H, Azimzai Y, Baughn MR, Tran B;
 XX
 DR WPI: 2000-549264/50.
 DR P-PSDB; AAY97285.
 XX
 XX New human lipid-associated proteins, nucleic acids, and antibodies,
 PT useful for diagnosis, treatment and prevention of e.g. cardiovascular
 PT disease
 XX
 PS Claim 4; Page 87-88; 93pp; English.
 XX
 CC Lipid-associated proteins (LIPAP) can be used for treating or
 CC preventing disorders associated with decreased expression of LIPAP,
 CC for screening for agonists or antagonists of LIPAP, and to raise
 CC specific antibodies. Antagonists and agonists of LIPAP are useful
 CC for treating diseases associated with reduced or increased levels of
 CC LIPAP, e.g. cardiovascular, neurological and gastrointestinal
 CC diseases and disorders of lipid metabolism. Fragments of the nucleic
 CC acid encoding LIPAP are useful for detection of full length coding
 CC sequences, in hybridization and/or amplification assays or for
 CC diagnosis or monitoring. Nucleotides encoding LIPAP are used
 CC to screen for compounds that specifically modify LIPAP expression,
 CC for recombinant production of LIPAP, in gene therapy, as a source of
 CC therapeutic antisense, triplex-forming, or ribozyme agents and for
 CC genomic mapping. Antibodies to the proteins are used for diagnosis
 CC and monitoring of LIPAP-associated disease by immunoassay, as
 CC antagonists, in competitive drug screens and for affinity
 CC purification of natural LIPAP.
 XX
 SQ Sequence 3705 BP; 1217 A; 743 C; 826 G; 919 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.71e-272 Length: 3705
 Score: 3927.00 Matches: 759
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-890-549-4 (1-759) x AAAS3826 (1-3705)
 QY 1 MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
 DB 137 ATGGAATCATCTCTTAAATAGCGCAATGACCTACTATCATTTGAGGGTAACAGCC 196
 QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
 DB 197 AAAGAACTTCTCTTGTCAACAGAACAAAGTCAATGGCTATTGTGGAATAATCTCCAAAG 256
 QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
 DB 257 TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAGAGAGAGTAAACCCGAAATCTC 316
 QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80

317 TCCAGCAGCTTTAGAAAGGGGACCCTGACTGTGTTAAAGAAAGTGGAGAACCCAGGG 376
 QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
 DB 377 CTGGAGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACACA 436
 QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
 DB 437 GCAGACCATCTCTCTGCTGAAGTCAAGCCACGCTGCTTCTGGAGCAAAAGTCAACCAA 496
 QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
 DB 497 GAAGAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAAAGCCCTCTGTTCAGGT 556
 QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
 DB 557 CGATATCCCCACATCAAGGACGCTGAGGATCTTAAGACCACTCAACAGAAAGTAAAAAA 616
 QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 DB 617 ATGGAAATTTGTCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAANAAC 676
 QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 DB 677 ACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAACAGGCTTAAGATGATG 736
 QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 DB 737 TTTGGAAGAGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 796
 QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
 DB 797 GGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCCAGGTGAG 856
 QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
 DB 857 TTGTCATCTTACATTTGACTTCGGAGAAAAATGAGAGTAGACGAGAAATCTGGAACTTCCA 916
 QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
 DB 917 CCGCTCTCAGAAACCTCTATAAAGATGCGAATGCCAAGTAGCCAGGACGTGTGTCCAAA 976
 QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
 DB 977 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGGTGGCGAAATCAAAAT 1036
 QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 DB 1037 CATAAAATGGAGCAAAAGGAGAATGTGCCCCAGGTCTCTGAGGTCTGCATCACCCTCAG 1096
 QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 DB 1097 GAAGGGGAAAAAGATTCTGCAAAATGAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAA 1156
 QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
 DB 1157 GATGACTCCCGTGACTCCCGAGTTAAGAGTAGGTTCAACAGCCCTGTCCATCCCAAGCCA 1216
 QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
 DB 1217 CTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAG 1276
 QY 381 LysPheGlnIleProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
 DB 1277 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGATGTTCAGAGAGAGAGTCTATCCAAATG 1336
 QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
 DB 1337 GAGCGTCTCTTGCGCAACCCAGCAGGTGTTTTCACATCAGCTGCTTCGTTCTCTCTATTGC 1396
 QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
 DB 1397 AACACAAACTCAGTCTAGGAACATATGATCTTTTACATGGAAAGATCTATTGTAAGCCT 1456

QY 441 HisPheAenGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1457 CACITCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGCTTTGGGCACAGACCA 1516
QY 461 HisLysAspLeuTtpAlaSerLysAsnGluGluLeuLeuGluAtpProAlaGln 480
DB 1517 CACAAGGATCTATGGGCAAGCAAAATGAACCAAGAGATTTTGGAGACACCGCCAG 1576
QY 481 LeuAlaAenAlaAaGgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1577 CTTGCAATGCAAGGGACACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTTAAG 1636
QY 501 ValGlyValLeuAlaAaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
DB 1637 GTGGGTGTCCTGGTCAAGTATGGAAGCAAGCCCTCTCTCAGCAGGAGAGCAAGAC 1696
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
DB 1697 AAGCCAGCTGAACCAAGAGAGCTGAGGATCCCTGGCCACCCCTCACTGAACCTTGAAGT 1756
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpPtcProGluAsp 560
DB 1757 TCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAGGCCCAATGGCTCTCTGAAGAC 1816
QY 561 GluLysSerLysProGluValProGluAspValAspLeuAspLysLysLysLeuArg 580
DB 1817 GAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTTAAGACGA 1876
QY 581 SerSerSerLysLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1877 TCTTCTTCACTGAAGGAAAGGCGGCCCATTCATCTAGCAGCTTCATTTCAAAGCACC 1936
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSerGlu 620
DB 1937 TCTGTCAGAGCCCAAAATCTGTCCTCCACCTATCAGGAAGGCTGAGCATGTTCAGAG 1996
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
DB 1997 CAGAGTGAAGAGTCTGTGGTGAAGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAG 2056
QY 641 AlaSerLysLysAenGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
DB 2057 GCITCTTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAATCTAAGGA 2116
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAenGluAenLeuVal 680
DB 2117 GAGACAGGGAAGAGAGTAAGAGAGTCAAGTTTGGAGATGAGATGAGATCTTGA 2176
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
DB 2177 GAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCAACAACAATCTCCACA 2236
QY 701 GluProLysSerLeuAenTtpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
DB 2237 GAACCCCAAGTCTCTGAATGGTGCAGTTTGTAGACAACACCTTGTGAGAATTCAC 2296
QY 721 ThrGlnAenGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
DB 2297 ACTCAGATCAGAAATCCCAAGATGTGGAACTCTGGAGGAGGAGATGTGTCAAGAGCTC 2356
QY 741 SerValGluGluGlnLysLysAaGsnArgTyrTyAspGluAspGluAspGluGlu 759
DB 2357 TCTGTGGAACAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2413

RESULT 4

AAC77143
ID AAC77143 standard; cDNA; 3711 BP.

XX AAC77143;

XX AC

DT 08-FEB-2001 (first entry)

XX

Human ORFX ORF2698 polynucleotide sequence SEQ ID NO:5395.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; anticholesterolemia; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB42934.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 4578-4580; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antirheumatic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 3711 BP; 1215 A; 747 C; 830 G; 919 T; 0 other;

Alignment Scores:

Pred. No.: 3,09e-271 Length: 3711
Score: 3909.50 Matches: 758
Percent Similarity: 99.74% Conservative: 0
Best Local Similarity: 99.74% Mismatches: 1
Query Match: 99.55% Indels: 1

DB: 21 Gaps: 1
US-09-890-549-4 (1-759) x AAC77143 (1-3711)
QY 1 MetGluSerSerProPheAsnArgGlnThrThrSerLeuSerLeuArgValThrAla 20
DB 142 ATGAATCATCTCCATTTAATAGACGCAATGACCTCACTATCATTTGAGGGTAACAGGC 201
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerLys 40
DB 202 AAGAAGCTTTCTTGTCAACAGACACAGCTATCGGCTATTGTGGAAATATTCTCCAAG 261
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
DB 262 TACCAGAAAGCAGCTGAAGAAACAAACATGGAGAGAGAGAGTAACACCCGAAATCTC 321
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
DB 322 TCCAGACACTTTAGAAGGGGACCCCTGACTGTGTGTTAAAGAAAGTGGGAGAACCCAGGG 381
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
DB 382 CTGGAGACAGAGTCTCACACAGACTCTCTACGGAGACAGCAGCAGTGGAGTATAGGCACAGA 441
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
DB 442 GCAGACCATCTCTCTGAAGTGACAGCCAGCTGCTTTGGAGCCAAAGCTGACCAA 501
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB 502 GAAGAACAAATCACCCAGACTCTAGACTCAGTCACTCTCTGAGGCCCTCTGTCAGGGT 561
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB 562 CGATATCCCCCATCATCAAGACCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 621
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
DB 622 ATGGAAATGTCTAGGAGATCCAGGCATGAAGTAGAAATATCAGAAATCAGTGAAAC 681
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB 682 ACAGATGCTTCGGGCAAAATACAGAAATATATATGTTCCGCTGAACAGGCTTAAGATGATG 741
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 742 TTGAGAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGGTCAAGT 801
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
DB 802 GGAAGAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAC 861
QY 241 LeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 862 TTGTCATCTTCACTTGAACCGAGAAATGAGAGTAGACGAAATCTGGAACTTCCA 921
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
DB 922 CGCTCTCAGAAACCTCTATAAGGATCGAATGCCAAGTACCGAGGAGCTGTGTCAAA 981
QY 281 GlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
DB 982 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAAT 1041
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB 1042 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTGAGGCTCGCATCACCCATCAG 1101
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1102 GAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCTGCGACTCGTTCACCCCTGCCGAA 1161
QY 341 AspAspSer---ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359

DB 1162 GNTAGTCCCCAGGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCCCAAG 1221
QY 360 ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMet 379
DB 1222 CCACTAAGTCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCTCCTCCCAAGCAATG 1281
QY 380 LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro 399
DB 1282 AAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAGAACACAGTCTATCCA 1341
QY 400 MetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyr 419
DB 1342 ATGGAGCTCTCTTGGCCACACAGCAGGTGTTTCAATCAGCTGCTTCCTGCTCTAT 1401
QY 420 CysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
DB 1402 TGCAACAACAACCTCAGTCTAGGAAACATATGCATCTTTTACATGGAAGATCTATTGTAAG 1461
QY 440 ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArg 459
DB 1462 CCTCACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGA 1521
QY 460 ProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAla 479
DB 1522 CCACACAGGATCTTATGGCAAGCAAAATGAAACGAAGAGATTTTGGAGAGACAGCC 1581
QY 480 GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAla 499
DB 1582 CAGCTTGCATAATGCAAGGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTGCT 1641
QY 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
DB 1642 AAGTGGGTGCTCTGGTGTGCAAGTATGGAAGCCAGAGCCCTCTCTCAGCAGGAGAAAGAA 1701
QY 520 AspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGly 539
DB 1702 GACAAGCAGCTGAACCAAGAGAGTGAGGATCGCTGGCCACCCCCACTGAACTTGGGA 1761
QY 540 SerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProProGlu 559
DB 1762 AGTTCAGAAAGTCTTGGAGAGAGGATCAAAATGTCAAAAGCCCAATGGCTCTCTGAA 1821
QY 560 AspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArg 579
DB 1822 GACGAAATCAGCAAGCCGAGGATTCCTGAGGATGTGATCTAGATCTGAAGAGAGCTAAGA 1881
QY 580 ArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSer 599
DB 1882 CGATCTTCTCACTGAAGGAAAGAGCCGCCCATTCCTGTAGCAGCTTCATTTTCAAGGC 1941
QY 600 ThrSerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMetSer 619
DB 1942 ACCTCTGTCAAGAGCCCAAAAGCTGTGCCACCTATCAGGAAAGGCTGGAGCATGTCA 2001
QY 620 GluGlnSerGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAla 639
DB 2002 GACACAGTCAAGAGCTCTGTGGTGGAAGAGTGCAGAAAGGAAACAAGTGGAAATGCC 2061
QY 640 LysAlaSerLysLysAsnGlnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
DB 2062 AAGCTTCTTAAAGAAATGGGAATGGGAAATAACACCTGGCAAAACAAAGAAATCTAAA 2121
QY 660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
DB 2122 GGAGACAGAGGAGAGAGAGTAAAGAGTCAATGTTTGGAGATGGAGATGAGAAATCTTT 2181
QY 680 ValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerPro 699
DB 2182 GTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCAACAAACAATCTCCA 2241
QY 700 GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719

Db 2242 CAAGAACCCCAAGTCTCTGTAATTGTCAGAGTTTGTAGACACACCTTTGTCGAGAAATTC 2301
QY 720 ThrThrGlnAnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGlu 739
Db 2302 ACTACTCAGAAATCCAGAGTGTGGAACCTCTGGAGGGAGAGTGGTCAAGAG 2361
QY 740 LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2362 CTCCTCTGTGGAAGAACAGATAAAGAGAAATCGGATATTATGATGAGGATGAGGATGAAGAG 2421
RESULT 5
AAF55697
ID AAF55697 standard; DNA; 3650 BP.
XX
AC AAF55697;
XX
DT 11-JUN-2001 (first entry)
XX
DE Human EPLIN (epithelial protein lost in neoplasm)-beta isoform DNA.
XX
KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer; ss.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
CDS 102..2384
FT /*tag= a
FT /transl_except= "(pos: 933..935, aa: Val)"
FT /transl_except= "(pos: 1131..1136, aa: Arg)"
FT /transl_except= "(pos: 1587..1589, aa: Asp)"
FT /transl_except= "(pos: 1635..1637, aa: Lys)"
FT /transl_except= "(pos: 2068..2070, aa: Ser)"
FT /transl_except= "(pos: 2076..2078, aa: Ser)"
FT /transl_except= "(pos: 2139..2141, aa: Ile)"
FT /transl_except= "(pos: 2196..2198, aa: Ser)"
FT /transl_except= "(pos: 2214..2216, aa: Ser)"
FT /product= "EPLIN (epithelial protein lost in neoplasm)"
XX
WO200118019-A1.
PN
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24689.
XX
PR 08-SEP-1999; 99US-0153024.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chang DD, Maul RS;
XX
XX WPI; 2001-244555/25.
DR P-PSDB; AAB67701.
XX
PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN -
XX
PS Claim 4; Page 43-44; 59pp; English.
XX
CC The present sequence encodes a human EPLIN (epithelial protein lost in
CC neoplasm)-beta isoform. The specification also describes EPLIN-alpha.
CC EPLIN is a tumour suppressor protein, whose expression is altered in
CC multiple common human tumour types. EPLIN nucleic acids and proteins are
CC used in screening assays to detect molecules that specifically bind to
CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
CC as agonist or antagonist of EPLIN, in particular molecules that affect
CC cell proliferation. Thus the assays are useful for screening molecules
CC with potential utility as anticancer drugs or lead compounds for drug
CC development. EPLIN nucleic acids, proteins are useful for detecting a
CC cell proliferative disorder in a subject. EPLIN polynucleotides are
CC useful in gene therapy techniques. EPLIN is useful as a marker that

CC can be diagnostically, prognostically and therapeutically used over
CC the course of a cell proliferative disorder associated with EPLIN.
XX
SQ Sequence 3650 BP; 1197 A; 733 C; 809 G; 911 T; 0 other;
Alignment Scores:
Pred. No.: 1.38e-269 Length: 3650
Score: 3872.50 Matches: 750
Percent Similarity: 99.08% Conservativity: 3
Best Local Similarity: 98.68% Mismatches: 6
Query Match: 98.61% Indels: 1
DB: 22 Gaps: 1
US-09-890-549-4 (1-759) x AAF55697 (1-3650)
QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 102 ATGGAAATCATCTCCATTTAATAGACGGAATGACCTCATTATGAGGGTAACAGCC 161
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 162 AAGAACAATCTCTTGTGTCACAAAGAACCAATCATCGGTATTGTGGAATATTCTCCAA 221
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 222 TACCAGAAAGCAGCTCAAGAAACAAACATCGAGAAAGAGAGAAATCAACCGAAATCTC 281
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 282 TCCAGACACTTTAGAAAGGGACCCCTGACTGTGTTTAAAGAAAGAGTGGAGAACCCAGG 341
QY 81 LeuGlyValGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 342 CTGGGAGCAGAGTCTCACACAGACTCTCTACGAAACACAGCAGCTAGATTAGGCACAGA 401
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValAlaAlaAspGln 120
Db 402 GCAGACCACTCTCTGCTGCTGAAGTGACAGCAGCCGCTGCTTCTGGAGCCAAAGCTGACCA 461
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 462 GAAGAACAAATCCACCCAGATCTAGACTCAGGTACCTCCTCAAGCCTCGTTCCAGGT 521
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 522 CGATATCCCCACATCAAGGACGGTGAAGATCTTAAAGACCACTCAACAGAAAGTAAAAA 581
QY 161 MetGluAsnCysLeuGlySerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 582 ATGGAATTTGCTTAGGAGAAATCCAGGCATGAGTAGAATAATCAGAAATCAGTGAATAAC 641
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 642 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATG 701
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 702 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCGGGGCCCAAGCCGAGTGCAGT 761
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
Db 762 GGAAGGAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCCCAGGTGAG 821
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 822 TTGTCACTTCTACATTGACTCGAGAGAAATAGAGTAGAGTAGACGAAATCTGGAACTTCCA 881
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 882 CGCTCTCAGAAACCTCTATAAGGATCGAATGCGCAAGTACCAAGCAGGAGCTGGCTCCAA 941
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300

942 CAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATT 1001
QY HisIysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
Db CATAAATGGAGCAAAAGGAGATGTCCCCCAAGTCTCGAGTCTGCATCAACCCATCAG 1061
QY GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db GAAGGGAAAGATTCTCGCAATGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1121
QY AspAspSer---ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359
Db GATGACTCCCGAGTGAAGTCCAGGTTAAGAGTGAGGTTCAACAGCGTGTCCATCCCAAG 1181
QY ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMet 379
Db CCCTAAGTCCAGATCCAGAGGCTCCAGGCTCTTCTGAAAGTCTCTCTCCCAAGCAATG 1241
QY LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro 399
Db AAGAGTTTCAGGCACCTGCAAGAGAGACCTGGTGGATGTGAGAGACAGTCTATCCA 1301
QY MetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyr 419
Db ATGAGCGTCTCTGGCCAAACAGCAGCAGTGTTCACATCAGCTGCTTCCGTTCTCTAT 1361
QY CysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
Db TGCACCAACAACATCAGTCTAGGACATATGTCATCTTACATGGAAGATCTATTGTAAG 1421
QY ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArg 459
Db CCTCAGTTCATCACTCTTAAATCTAAGGGCACTATGATGAAGGCTTGGGCACAGA 1481
QY ProHisLysAspLeuTrpAlaSerLysAsnGluGluLysLeuGluArgProAla 479
Db CCACACAGGATCTATGGCAACCAAAATGAAACGAGAGATTTTGAGAGACAGCC 1541
QY GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAla 499
Db CAGCTTGCAAAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAAAATGCCCTATGCT 1601
QY LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGlu 519
Db AAGTGGGTCTCTGGCTGCAAGTATGGAAGCCAGGGCTCTCTCTCAGCAGGAGAGAA 1661
QY AspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGly 539
Db GACAGCCAGCTGAACCAAGAGCTGAGATCGCTGGCCCTCCACCCCTGAACTTGA 1721
QY SerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProGlu 559
Db AGTTCAGGAAGTCCCTTGAGGAAGGATCAAAATGTCAAAGCCCAATGGCTCTGNA 1781
QY AspGluLysSerLysProGluValProGluAspValAspLeuAspLysLysLeuArg 579
Db GAGCAATCAGCAAGCCCGAAGTTCCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGA 1841
QY ArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSer 599
Db CGATCTTCTTCACTGAGGAAGAAGCCGCCCATTCACGTGACAGCTTCATTTCAAGC 1901
QY ThrSerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSer 619
Db ACCTCTGTCAAGAGCCCAAAATGCTGCCACCTATCAGAAAGGCTGGAGCATGCA 1961
QY GluGlnSerGluLysValGlyArgValAlaGluArgLysGlnValGluAsnAla 639
Db GAGCAGATGAAGATCTCTGGTGGAGAGTTCAGAAAGGAAACAAAGTGGAAATGCC 2021
QY LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
Db AAGGCTTCTAAGAGAAATGGGAATGGGAAAAACCACTGGCAAAACAAAGAAATTTAA 2081

660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
Db GGAGAGACAGGAGAGAGAGAGTAAAGCAAGTCTATAGTTTGGAGATGGAGATTTT 2141
QY ValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnSerPro 699
Db GTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACCAAAATTTCCA 2201
QY GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719
Db CAAGAACCCAGTTCGAATTTGTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 2261
QY ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValLysGlu 739
Db ACTACTCAGATCAGAAATCCAGGATGTGGAACCTTTGGAGGGAGAGTGGTCAAGAG 2321
QY LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db CTCTCTGGAGAACAGATTAAGAGAAATCGGTATTATGATGAGGATCAGGATCAGGATCAGGAT 2381
RESULT 6
AAF55696
ID AAF55696 standard; DNA; 3543 BP.
XX
AC AAF55696;
XX
DT 11-JUN-2001 (first entry)
XX
DE Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform DNA.
XX
KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 474..2276
FT /cds= a
FT /transl_except= "(pos: 825..827, aa: Val)"
FT /transl_except= "(pos: 1479..1481, aa: Asp)"
FT /transl_except= "(pos: 1627..1629, aa: Lys)"
FT /transl_except= "(pos: 1860..1861, aa: Ser)"
FT /transl_except= "(pos: 1968..1970, aa: Ser)"
FT /transl_except= "(pos: 2031..2033, aa: Ile)"
FT /transl_except= "(pos: 2088..2090, aa: Ser)"
FT /transl_except= "(pos: 2106..2108, aa: Ser)"
FT /product= "EPLIN (epithelial protein lost in neoplasm)"
XX
WO200118019-A1.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24689.
XX
PR 08-SEP-1999; 99US-0153024.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chang DD, Maul RS;
XX
DR WPI; 2001-244555/25.
XX P-PSDB; AAB67700.
XX
PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN -
PS Claim 4; Page 42-43; 59pp; English.
XX
CC The present sequence encodes a human EPLIN (epithelial protein lost in
CC neoplasm)-alpha isoform. The specification also describes EPLIN-beta.

CC EPLIN is a tumour suppressor protein, whose expression is altered in
 CC multiple common human tumour types. EPLIN nucleic acids and proteins are
 CC used in screening assays to detect molecules that specifically bind to
 CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
 CC as agonist or antagonist of EPLIN, in particular molecules that affect
 CC cell proliferation. Thus the assays are useful for screening molecules
 CC with potential utility as anticancer drugs or lead compounds for drug
 CC development. EPLIN nucleic acids, proteins are useful for detecting a
 CC cell proliferative disorder in a subject. EPLIN polynucleotides are
 CC useful in gene therapy techniques. EPLIN is useful as a marker that
 CC can be diagnostically, prognostically and therapeutically used over
 CC the course of a cell proliferative disorder associated with EPLIN.
 XX

SQ Sequence 3543 BP; 1152 A; 711 C; 771 G; 909 T; 0 other;

Alignment Scores:

Pred. No.: 2,866-249 Length: 3543
 Score: 3603.50 Matches: 695
 Percent Similarity: 99.01% Conservative: 3
 Best Local Similarity: 98.58% Mismatches: 6
 Query Match: 91.76% Indels: 1
 DB: 22 Gaps: 1

US-09-890-549-4 (1-759) x AAF55696 (1-3543)

QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
 DB 159 AACACCGAATCTCTCCACACCTTTTAGAAGGGGACCCCTGACTGTGTTAAAGAGAAG 218
 QY 76 TrpGluAsnProGlyLeuGluValAlaGluSerHisThrAspSerLeuArgAsnSerSerThr 95
 DB 219 TGGAGAACCCAGGCTGGAGCAGAGTCTCACACAGACTCTTACGGAACAGCAGCACT 278
 QY 96 GluLeuArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly 115
 DB 279 GAGATTAGGCACAGCAGACCATCTCTCTGCTAGGTGACCAAGCCAGCTGCTTCTGGA 338
 QY 116 AlalysAlaAspGlnGluGluLeuHisProArgSerArgLeuArgSerProProGlu 135
 DB 339 GCCAAAGCTGACCAAGAGAAGAACAAATCCACCCAGATCTAGACTCAGTCCACTCTCTGAA 398
 QY 136 AlaLeuValGlnGlyArgTyProHisLysAspGlyGluAspLeuLysAspHisSer 155
 DB 399 GCCCTCGTTTCAGGTGCGATATCCCCACATCAAGGACCGTGAGGATCTTAAAGACCACTCA 458
 QY 156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
 DB 459 ACAGAAAGCTAATAAATGGAAAATTGCTAGCAGAAATCCAGGCATGAAGTAGAATAATCA 518
 QY 176 GluLeuSerGluAsnThrAspAlaSerGlyLysLysLeuLysTyAsnValProLeuAsn 195
 DB 519 GAAATCAGTGAAGAACACAGATGCTTCGGCAAAATAGAGAAATATAATGTTCCGCTCAAC 578
 QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuArgAlaGln 215
 DB 579 AGGCTTAAGATGATGTTTGAAGAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAA 638
 QY 216 SerArgSerAlaSerGlyArgLysLysSerGluAsnSerTySerLeuAspAspLeuGlu 235
 DB 639 AGCCGAAGTGAAGTGAAGAAATCTCTGAAAACAGCTATTCTTAGATGACCTGGAA 698
 QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
 DB 699 ATAGGCCAGGTGAGTTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGACGA 758
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerLysAspArgMetAlaLysTyGln 275
 DB 759 AATCTGAAGACTTCACGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAGTACCAG 818
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyThrAsnGluLeuLysAlaSerGly 295
 DB 819 GCAGCTCGTCCAAACCAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 878

QY 296 GlyGluLeuLysLysLysHisLysMetGluGlnLysGluAsnValProProGlyProGluVal 315
 DB 879 GGGCAATCAAAATTCATAAATGGAGCAAAAGAGATGTCCTCCCCAGGTCTCTGAGGTC 938
 QY 316 CysIleThrHisGlnGluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaValArg 335
 DB 939 TGCATCACCCTCAGGAAGGGGAAAGATTCTGCAATGAGAATAGCTTGGCAGTCCGT 998
 QY 336 SerThrProAlaGluAspAspSer---ArgAspSerGlnValLysSerGluValGlnGln 354
 DB 999 TCCACCCCTGCCAAGATCACTCCCGAGTGACTCCAGGTTAAGAGTAGGTTCAACAG 1058
 QY 355 ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer 374
 DB 1059 CTTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCT 1118
 QY 375 ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln 394
 DB 1119 CTTCCCAAGCATGAAGAGATTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCA 1178
 QY 395 LysThrValTyProMetGluArgLeuAlaAsnGlnValPheHisLysSerCys 414
 DB 1179 AAGACAGTCTATCCAAATGGAGGCTCTCTTGGCCAACCCAGCAGGTGTTTCCATCAGCTGC 1238
 QY 415 PheArgCysSerTyProCysAsnAsnLysLeuSerLeuGlyThrTyAlaSerLeuHisGly 434
 DB 1239 TTCCTGTGCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCACTCTTTACATGGA 1298
 QY 435 ArgIleTyCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyAspGlu 454
 DB 1299 AGAATCTATGTGAAGCCTCCTCAATCAACTCTTAAATCTAAGGGCACTATGATGAA 1358
 QY 455 GlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluLe 474
 DB 1359 GCGTTTGGGCACAGACCCACACAGGATCTATGGGAAGCAAAATGAAACCAAGAGATT 1418
 QY 475 LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGlu 494
 DB 1419 TTGGAGAGACCCAGCCAGCTTCCAAATGCAAGGGAGACCCCTCAGCCAGGGGTAGAA 1478
 QY 495 AspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSer 514
 DB 1479 AATGCCCTTATTGCTAAGGTGGTGTCTCTGGTGTCAAGTATGGAAGCCAGGCCCTCTCT 1538
 QY 515 GlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgLysAlaTrpProPro 534
 DB 1539 CAGCAGGAGAGAGAACAGCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCC 1598
 QY 535 ProThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysLysMetSerLysPro 554
 DB 1599 CCCACTGAACCTTGGAGTTTCAGAAAGTCCCTTGGAGGAAGGATCAAAATGTCAAAGCCC 1658
 QY 555 LysTrpProProGluAspGluLysSerLysProGluValProGluAspValAspLeuAsp 574
 DB 1659 AAATGGCTCTCTGAAGCAAAATCAGCAAGCCGGAAGTTCTCTGAGGATGTCGATCTAGAT 1718
 QY 575 LeuLysLysLeuArgArgSerSerSerLysLysGluArgSerArgProPheThrValAla 594
 DB 1719 CTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAACCGCCCATTCACCTAGCA 1778
 QY 595 AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProLysLeuArgLys 614
 DB 1779 GCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1838
 QY 615 GlyTrpSerMetSerGluGlnSerGluGlyGlyArgValAlaGluArgLys 634
 DB 1839 GCTCGACATGTCAGACAGATGAAGANTCTGTGGGTGGAAGAGTTGCAAGAAAGGAAA 1898
 QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGln 654
 DB 1899 CAAAGTGAAGAAATGCCAAGGCTTCTAAGAGAAATGGAATGTTGGGAAAACCACTGGCAA 1958
 QY 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluVet 674

Db 1959 AACAAAGAAATTTAAAGGAGAGAGAGGAGAGAGTAAGGAAGTCTATGTTGGAGATG 2018
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeu 694
Db 2019 GAGAAATGAGAAATTTGTAGAAATGTTGTCAGACTCCGATCAAGATGATAACAGCTTCCIC 2078
Qy 695 LysGlnGlnSerProGlnGluProLysSerLeuSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2079 AAACAAATTTCCACAAAGAACCCAGATTTTGAATGGTCGAGTTTGTAGACAAACACC 2138
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2139 TTTGCTGAGCAATTCACACTCAGATCAGAAATCCAGAGATGTGAACTTTGGGAGGGA 2198
Qy 735 GluValValLysGluLeuSerValGluGluGlnLeuLysArgAsnArgTyrTrpAspGlu 754
Db 2199 GAAGTGTGTCAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG 2258
Qy 755 AspGluAspGluGlu 759
Db 2259 GATGAGATGAAGAG 2273

RESULT 7
AAS72746
ID AAS72746 standard; cDNA; 3465 BP.
AC AAS72746;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #8550.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG08559.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 8550; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3465 BP; 1112 A; 706 C; 773 G; 874 T; 0 other;

Alignment Scores:
Pred. No.: 1.98e-226 Length: 3465
Score: 3285.50 Matches: 688
Percent Similarity: 94.7% Conservative: 1
Best Local Similarity: 94.6% Mismatches: 15
Query Match: 83.6% Indels: 23
DB: 23 Gaps: 5

US-09-890-549-4 (1-759) x AAS72746 (1-3465)

Qy 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLys 75
Db 4 AACACCGAAAAATCTCTCCAGCACTTTAGAAAAGGGACCTGACTGTGTTAAAGAAGAAG 63
Qy 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThr 95
Db 64 TGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGAGCACT 123
Qy 96 GluLeuArgHisArgAlaAspHisProAlaGluValThrSerHisAlaAlaSerGly 115
Db 124 GAGATTAGGCACAGACAGACCATCTCTCTGCTGAAAGTGACAAAGCCACGCTTCTTCTGA 183
Qy 116 AlalysAlaAspGlnGluGlnHisProArgSerArgLeuArgSerProProGlu 135
Db 184 GCCAAAGCTGACCAAGAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAA 243
Qy 136 AlaleuValGlnGlyArgTyrProHisLysAspGlyGluAspLeuLysAspHisSer 155
Db 244 GCCCTCGTTTCAGGTCGAT 303
Qy 156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
Db 304 ACAGAAAGTAAAAAATGGAAAAATGTTCTAGGAAATCCAGGCATGAAGTAGAAAAATCA 363
Qy 176 GluLeuSerGluAsnThrAspAlaSerGlyLysLysLysGlyLysValProLeuAsn 195
Db 364 GAAATCAGTGAAPACACAGATGCTTCGGGCAAAATAGAGAAATATATATATATATAT 423
Qy 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuAlaGln 215
Db 424 AGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTTAAGATTCTCCGGGCCCAA 483
Qy 216 SerArgSerAlaSerGlyArgLysLysLysLysLysLysLysLysLysLysLysLys 235
Db 484 AGCCGAAGTCAAGTGAAGAGAGATCTCTGAAACACAGCTATTCTCTAGATGACTCGAA 543
Qy 236 IleGlyProGlyGlnLeuSerSerThr-PheAspSerGlyLysAsnGluSerArgAr 255
Db 544 ATAGGCCACAGTCAGTTGTCTATCTTACATTTTGACTCGGAAAAAATAGAGATAGACG 603
Qy 255 GAsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlalysTrpG 275
Db 604 AAATCTGGAATCTCCAGCCCTCTCAGAACCTCTATAAAGATCGAATGCGCAAGTACCA 663
Qy 275 nAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerG 295
Db 664 GGCAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAATCCAGCGG 723
Qy 295 YGlyGluLeuLysLysLysMetGluGlnLysGluAsnValProProGlyProGluVa 315
Db 724 TGGCAAAATCAAAATTCATAAAACCGGCAAGGAGAAATGTGCCCCAGGCTCTGAGGT 783

QY 315 lCysIleThrHis-GlnGluGlyGluIysIleSerAlaAsnGluAsnSerLeuAlaVala 335
Db 784 CTGCATCACCCATTGAGGAGGGAAGATTTCTGCAATAGAGATAGCCTGCGAGTCC 843
QY 335 rGSerThrProAlaGluAspSer---ArgAspSerGlnValIysSerGluValGlnG 354
Db 844 GTTCCACCCCTGCGCAGAGTACCTCCAGGTGAGTCCAGGTTAAGAGTGAGGTTCAC 903
QY 354 lnProValHisProLysProLysProLysProLysProLysProLysProLysPro 374
Db 904 AGCCTGTCCATCCCAAGCCATAGTCCAGATTCAGAGCCTCCAGTCTTCTGAAAGT 963
QY 374 erProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysG 394
Db 964 CTCCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAGTGC 1023
QY 394 lnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerC 414
Db 1024 AGAAGACAGTCTATCCCAATGAGAGCTCTCTTGGCCAAACCAAGAGTCTTTACATCAGCT 1083
QY 414 ysPheArgCysSerTyrcysAsnAsn-LysLeuSerLeuGlyThrTyrrAlaSerLeuHis 433
Db 1084 GCTTCCGTGTCTCTATTGCAACCAACAACTCAGTCTAGAACATATGATCTTTACAT 1143
QY 434 GlyArgIleTyrcysLysProHisPheAsnGlnLeuPheLysSer-LysGly-AsnTyrrA 453
Db 1144 GGAAGATCTATTGAGCCTCACTTCAATCACTTTAAATCTTAAGGGCAAACTATG 1203
QY 453 spGluGlyPheGlyHisArgProHisLysAspLeu--TrpAlaSerLysAsnGluAsn-G 472
Db 1204 ATGAAGGCTTTGGGCACAGACACACAAAGGATCTTATTGGGCAAGCAAAATGAAAAAG 1263
QY 472 luGluIleLeuGluArgProAla-GlnLeuAlaAsnAlaArgGluThrProHisSerPro 491
Db 1264 AGGAGATTTTGGAGAGACCCAGCCCGCTGCAATGCAAGGGAGACCCCTCACAGCCCA 1323
QY 492 GlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLys 511
Db 1324 GGGGTAGAAGATGCCCCATTGCTAAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1383
QY 512 AlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAla 531
Db 1384 GCCTCCTCTCAGCAGGAG 1443
QY 532 TrpProProThrGluLeuGlySerGlySerGlySerGlySerGlySerGlySerGlyMet 551
Db 1444 TGGCCACCCCCCTGAACTGGAAGTTCAGGAGTGCCTTGGAGGAGGAGATCAAAATG 1503
QY 552 SerLysProLysTrpProGluAspGluLysSerLysProGluValProGluAspVal 571
Db 1504 TCAGAGCCCAATGGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563
QY 572 AspLeuAspLeuLysLysLeuArgArgSerSerSerSerSerSerSerSerSerSerSer 591
Db 1564 GATCTAGATCTGAAGAAGCTAAGACGATCTTCTACTGAAGGAGAGAGAGAGAGAGAGAG 1623
QY 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
Db 1624 ACTGTAGACCTTCAATTCAAAGCACCCTCTGTCAAGAGCCCAAAATCTGTCTCCACCT 1683
QY 612 -IleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAl 631
Db 1684 TATCAGAAAGGCTGGAGAGATGTCAGAGCAGAGATGAGAGTCTGTGGGTGGAAGATGTC 1743
QY 631 aGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh 651
Db 1744 AGAAGAGAAACAAGTGGAAATGCCAGGCTTCTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
QY 651 rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLys-GluGly-His 670
Db 1804 AACCTGGCAAAACAAGAACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCCAT 1863

QY 671 SerLeuGluMetGluAsnGluAsnLeuVal--GluAsnGlyAlaAspSerAsp---GluA 689
Db 1864 AGTTTCAGATGAGATGAGATGAGATCTTTGTAGAAAATGCTCCAGACTCCCGATTGAAG 1923
QY 689 sPAspAsnSerPhe---LeuLysGlnGlnSer-Pro-GlnGluProLysSerLeu---As 706
Db 1924 ATGATAACCAAGCTTCTCCCAAAACAACATCTTCCACCAAGAACCAAGCTCTTCTTGAAT 1983
QY 706 nTrpSer-SerPheValAspAsn---ThrPheAlaGluGluPheThrThr-GlnAsnGln 724
Db 1984 TTGGTCCGAGTTTGTAGACAACCACTTTTCTGAGGAATTCATCTACTCCAGAAATCAG 2043
QY 725 LysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGlu 744
Db 2044 AATCCAGAGTGGAACTCTGGAGGGAGAGAGTGGTCAAGAGAGCTCTCTGTGGAGAA 2103
QY 745 GlnIleLysArgAsnArgTyrrAspGluAspGluAspGluGlu 759
Db 2104 CAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2148
RESULT 8
ABI98973
ID ABI98973 standard; cDNA; 1754 BP.
XX
AC ABI98973;
XX
DT 21-FEB-2002 (first entry)
XX
DE Human cancer suppressor protein PP624 encoding cDNA.
XX
KW Human; cancer suppressor, disease; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key
CDS 444..1349
FT /*tag= a
FT /transl_except= (pos:104..106,aa:Xaa)
FT /transl_except= (pos:561..563,aa:Xaa)
FT /transl_except= (pos:570..572,aa:Xaa)
FT /product= "PP624"
FT /*note= "Xaa = unknown"
XX
PN CN1313297-A.
XX
PD 19-SEP-2001.
XX
PF 09-MAR-2000; 2000CN-0111948.
XX
PR 09-MAR-2000; 2000CN-0111948.
XX
PA (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI Gu J, Yang S;
XX
DR WPI; 2002-042185/06.
P-PSDB; ABB56420.
XX
PT Human protein able to suppress growth of cancer cells and its coding
sequence -
XX
PS Claim 5; Page 21-22 Disclosure; 37pp; Chinese.
XX
CC The invention relates to novel human proteins (ABB56417-ABB56425) with
cancer suppressing function, the encoding polynucleotides
CC (ABI98970-ABI98978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
XX
SQ Sequence 1754 BP; 586 A; 377 C; 426 G; 361 T; 4 other;
Alignment Scores:

Pred. No.:	3,23e-183	Length:	1754
Score:	2679.00	Matches:	523
Percent Similarity:	98.87%	Conservative:	0
Best Local Similarity:	98.87%	Mismatches:	5
Query Match:	68.22%	Indels:	2
DB:	24	Gaps:	0
US-09-890-549-4 (1-759) x ABI98973 (1-1754)			

Qy	232	AspAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluIysAsn	251
Db	2	GATGACCTGGAATAGGCCAGGTGAGTTGTTCATCTTCTACATTGACTCGAGAAAT	61
Qy	252	GluSerArgArgAsnLeuGluLeuProArgLeuSerGluThrSerIleIysAspArgMet	271
Db	62	GAGAGTAGACGAAATCTGGAACTTCACGCCCTCTCAGAAACCTCTATAAAGGATCGAATG	121
Qy	272	AlaLysTyrGlnAlaAlaValSerIysGlnSerSerSerThrAsnTyrThrAsnGluLeu	291
Db	132	GCCAAAGTACCAGCAGCTGTGTCCAAACAAAGCAGCTCAACCACTATACAAATGAGCTG	181
Qy	292	LysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProPro	311
Db	182	AAAGCCAGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGAGAAATGTGCCCCA	241
Qy	312	GlyProGluValCysIleThrHisGlnGluGlyLysIleSerAlaAsnGluAsnSer	331
Db	242	GGTCTGAGGCTGCATCACCCATCAGGAAGGGGAAAGATTCTGCAAATGAGAAATAGC	301
Qy	332	LeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValIysSerGlu	351
Db	302	CTGCAGTCCGTTCCACCCCTGCCAAGATGACTCCCGTACTCCAGGTTAAAGAGTGCAG	361
Qy	352	ValGlnGlnProValHisProIysProIeuSerProAsp-SerArgAlaSerSerLeuSe	371
Db	362	GTTCAAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATCTCCAGAGCTCCAGTCTTTC	421
Qy	371	rGluSe-SerProProLysAlaMetLysLysPheGlnAlaProAlaAArgGluThrCysVa	391
Db	422	TGAAAGTTCCTCCCAAGCAATGAAGAGTTTCAGGACCTCGCAAGAGAGACTCGGT	481
Qy	391	lGluCysGlnIysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHi	411
Db	482	GGAATGTCAGAAGACAGTCTATRCAATGGAGGGTCTCTTGCCCAACAGCAGGTTTCA	541
Qy	411	sIleSerCysPheArgCysSerTyrCysAsnLysLeuSerLeuGlyThrTyrAlaIase	431
Db	542	CATCAGCTGCTTRCGTTGCTRCTATTGCAKCAACAAACTCAGTCTAGGAACATATGCATC	601
Qy	431	rLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnIeuPheIysSerLysGlyAs	451
Db	602	TTTACATGGAGAAATCTATTGTAGGCTCACTTCAATCAACTCTTTAAATCTAAGGGCAA	661
Qy	451	nTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerIysAsnGluAs	471
Db	662	CTATGATGAGGCTTTGGGCACAGCCACACAGGATCTATGGGCACAGCAAAAATGAAA	721
Qy	471	nGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerPr	491
Db	722	CGAAGAGATTTTGAGAGACACAGCCAGCTTGCAATGCGAAGGAGACCCCTCTACAGCCC	781
Qy	491	oGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLy	511
Db	782	AGGGGTAGAGATGCCCTATTGCTAAGTGGGTGTCTCGCTGCAAGTATGGAAGCCAA	841
Qy	511	sAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrIlysLysLeuArgIleAl	531
Db	842	GGCCTCCTCTCACGAGAGAGAAGAACAAAGCCAGCTGAAACCAAGAAAGCTGAGGATCG	901
Qy	531	aTrpProProThrGlnLeuGlySerSerGlySerAlaLeuGluClyIleLysMe	551
Db	902	CTGGCCACCCCACTGAATCTGGAGTTTCAGGAAGTGCTTCGAGAAAGGATCAAAAT	961

Qy	551	tSerLysProLysTrpProProGluAspGluIleSerLysProGluValProGluAspVa	571
Db	962	GTCAAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCGGAAGTTCCTGAGGATGT	1021
Qy	571	lAspLeuAspIleLysLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPh	591
Db	1022	CGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCTCAGAGGAAGAAAGCGCCCAT	1081
Qy	591	eThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPr	611
Db	1082	CACTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACCTGTGTCCCCACC	1141
Qy	611	oileArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAla	631
Db	1142	TATCAGGAAGGCGTGGAGCATGTCTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGC	1201
Qy	631	aGluArgLysGlnValGluAlaAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh	651
Db	1202	AGAAGGAACAAGTGGAAATCCCAAGGCTTCTAAGAAGAATGGGAAATGTGGGAAAAAC	1261
Qy	651	rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisE	671
Db	1262	AACCTGGCAAAACAAAGAATCTAAAGGAGCA-GGGAGAGAAGTAAAGAAAGTCTATG	1320
Qy	671	rLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsAs	691
Db	1321	TTTGGAGATCGAGAATCAGAATCTTGTAGAAAATGTGCAGACTCCGATGAAGATGATAA	1380
Qy	691	nSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheVa	711
Db	1381	CAGCTTCTCTCAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCAGTTTGT	1440
Qy	711	lAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLe	731
Db	1441	AGACAACACCTTTGCTGAAGAATTCTACTCAGAATCAGAAATCCAGGATGTGGAAC	1500
Qy	731	uTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTy	751
Db	1501	CTGGAGGGGAGAAGTGTCAAAGAGCTCTCTGTGGAGAAACAGATAAAGAAATCGGTA	1560
Qy	751	rTyAspGluAspGluAspGluGlu	759
Db	1561	TTATGATGAGGATGAGGATGAAGAG	1585
RESULT 9			
AA159955			
ID	AA159955 standard; cDNA; 1713 bp.		
XX	AA159955;		
AC			
XX			
DT	22-OCT-2001 (first entry)		
XX	Human polynucleotide SEQ ID NO 3944.		
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	anotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
XX	Leukaemia; ss.		
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		

PR 03-AUG-2000: 2000US-0653450.
PR 14-SEP-2000: 2000US-0682191.
PR 19-OCT-2000: 2000US-0693036.
PR 29-NOV-2000: 2000US-0727364.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM40799.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 3944; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1713 BP; 549 A; 401 C; 417 G; 346 T; 0 other;

Alignment Scores:
Pred. No.: 5,03e-173 Length: 1713
Score: 2537.00 Matches: 503
Percent Similarity: 95.87% Conservative: 8
Best Local Similarity: 94.37% Mismatches: 20
Query Match: 64.60% Indels: 4
DB: 22 Gaps: 0

US-09-890-549-4 (1-759) x AAI59955 (1-1713)

QY 1 MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 121 ATGGAATCATCTCCATTATATAGACGGCAATGGACCTCACTATCATTGAGGGTAACAGCC 180
QY 21 LysGluLeuSerLeuValLeuAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 181 AAGAACTTCTCTTGTCAACAAGACAGTATCGCTGCTATTTGTGAAATATTTCTCCAG 240
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 241 TACCAGAAACACGCTGAAGAAACAAACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysLysLysLysLysLys 80
Db 301 TCCAGACACTTTAGAAAGGGGACCTGACTGTGTTAAAGAGAGAGTGGGAGAACCCAGGG 360
QY 81 LeuGluValAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 361 CTGGGAGCAGAGTCTCACACAGACTCTTACCGAAGCAGAGCAGACTGAGATTAGGCACAGA 420
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValLysAlaAspGln 120
Db 421 GCAGACCATCTCTGCTGAAGTGAACAGCCAGCTGCTTCTGGAGCCAAAGCTCAGCAA 480
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140

Db 481 GAAGAAACAAATCCACCCAGATCTAGACTCAGGCACCTCTCTGAAGCCCTCGTTTCAGGCT 540
QY 141 ArgTyrProHisLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 541 CGATATCCCCCACCATCAAGGACGGTGGAGTCTTAAAGACCACTCAACAGAAAGTAAAAA 600
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluAsn 180
Db 601 ATGGAAATTTGTCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAGATCAGTGAAC 660
QY 181 ThrAspAlaSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 200
Db 661 ACAGATGCTTCGGGCAAAATAGAAAAATATATGTTCCGCTGAAACAGGCTTAAAGATGATG 720
QY 201 PheGluLysGlyGluProThrGlnThrLysLysLysLysLysLysLysLysLysLysLys 220
Db 721 TTTGAGAAAGGTGAACCACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 780
QY 221 GlyArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 240
Db 781 GGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAG 840
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgAtgAsnLeuGluLeuPro 260
Db 841 TTTGTCATCTTCTACATTTGACTCGAGAAATAATGAGATGACGAAATTTTGGAACTTCCA 900
QY 261 ArgLeuSerGluThrSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 280
Db 901 CGCCTCTCAGAAACCTCTATGAAGGATCGAATGGCAAGTAGTACCAGGACGTGTGTCCAA 960
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysLysLysLysLysLysLysLysLys 300
Db 961 CAAAGCACTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGGCAAAATCAAAAT 1020
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 1021 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGTCTGATCACCACATCAG 1080
QY 321 GluGlyGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 340
Db 1081 GAAGGGGAAAGATTTCTTGCAAAATGAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCAA 1140
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1141 GATGACTCCCGTGACTCCCGAGTTAAGAGTGAGGTTCACAGCTGTCCATCCCAAGCCA 1200
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1201 CTAAATCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAG 1260
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1261 AAGTTTCAGGCACCTGCAAGAGAGACCTGCTGGAATGTGAGAGAGAGTCTATCCCAATG 1320
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisLysSerCysPheArgCysSerTyrCys 420
Db 1321 GAGCGTCTCTTGGCCCAACCCAGCAGTGTTCACATCAGTGTCTCCGTTCTCTATTGC 1380
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgLysLysLysPro 440
Db 1381 AACAAACAACTCAGCTAGGAACATATGCAATCTTTACATGGAAGTCTGAGAGAGTCTAT 1440
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1441 CACTTCAATCACTCTTTAAATCTAAGGCACTATGATGAAGGCTTTGGGAGAGACCA 1500
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluLeuGluArgProAlaGln 480
Db 1501 CACAAGGATCTATGGGCAACAAAAATGAACCGAGGGGTTTGGGAGAGACCAAGC-CAA 1559
QY 481 LeuAlaAsnAlaArgGluThrProHis-SerProGlyValGluAsp-AlaProIleAlaL 500

Db 1560 CTTTGAATAATCAAGGAGAGCCCTCAAAAGCCAGGGGTGAAGATTGCCCTTCTGCTA 1619
Qy 500 ysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlnA 520
Db 1620 AGGTGGTGTCTCGCTCGAATTAATGAGCCCAAGGCTCTCTTTCACGAGAGAAGGAGG 1679
Qy 520 sPLysProAlaGluThrLysLysLeuArgIleAla 531
Db 1680 ACAAGCCAGCTGGACCCG-AAACTTAGGTGCGCC 1713

RESULT 10

ABK35261
ID ABK35261 standard; cDNA; 2749 BP.

AC ABK35261;

XX 08-MAY-2002 (first entry)

XX Human cDNA encoding secreted protein #399.

XX Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
XX viral infection; bacterial infection; fungal infection; diabetes; asthma;
XX autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
XX autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
XX Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
XX coagulation disorder; inflammatory disorder; Crohn's disease; incision;
XX tissue regeneration; wound healing; burn; haematopoiesis;
XX myeloid cell deficiency; lymphoid cell deficiency.

XX Homo sapiens.

XX W0200177288-A2.

XX 29-MAR-2001; 2001WO-US10224.

XX 18-OCT-2001.

XX 06-APR-2000; 2000US-195582P.

XX (GEMY) GENETICS INST INC.

XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;

XX WPI; 2002-179321/23.

XX Five hundred and ninety two polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders

XX Claim 1; Page 279-280; 372pp; English.

XX The invention relates to 592 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins. The polynucleotides can be used as probes for the
CC identification and isolation of full length cDNA and genomic DNA. The
CC polynucleotides and proteins can also be used as nutritional supplements.
CC The proteins are useful in the treatment of various immune deficiencies
CC and disorders such as viral infections, bacterial infections, fungal
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC and conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
XX Sequences ABK34863-ABK34854 represent polynucleotides of the invention.

XX Sequence 2749 BP; 885 A; 535 C; 593 G; 736 T; 0 other;

Alignment Scores:

Pred. No.: 1.27e-172 Length: 2749
Score: 2535.00 Matches: 490
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 64.55% Indels: 1
DB: 24 Gaps: 0

US-09-890-549-4 (1-759) x ABK35261 (1-2749)

Qy 269 AspArgMetAlaLysTyrGlnAlaAlaValSerLysGlnSerSerThrAsnTyrThr 288
Db 1 GATCGAATGGCCAGTACAGGACGTGTGTCCAAACCAAGCAGCTCAACCACTATACA 60
Qy 289 AsnGluLeuLysAlaSerGlyGlyGluIleLysLysLysMetGluGlnLysGluAsn 308
Db 61 AATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAATGGAGCA-AAGAGAAAT 119
Qy 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysLysSerAlaAsn 328
Db 120 GTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTACAGGAGGGGAAAGATTTCTGCAAT 179
Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
Db 180 GAGAAATAGCTGGCAGTCTCCACCCCTGCCAAGATGACTCCCGTGACTCCCGAGTT 239
Qy 349 LysSerGluValGlnGlnProValHisProLysProLysProLysProAspSerArgAlaSer 368
Db 240 AAGAGTGAAGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCC 299
Qy 369 SerLeuSerGluSerSerProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 300 AGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 359
Qy 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 360 ACCTGCGTGGAAATGTCAAGACAGAGTCTATCCATGGAGCGTCTCTTGCCCAACACGAGAG 419
Qy 409 ValPheHisLysSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428
Db 420 GTGTTCACATCAGCTGCTTCCGTGTCTCTATTGCAACACAACTCAGTCTAGGAACA 479
Qy 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 480 TATGCATCTTTACATGGAAGAAATCTATTGTAAGCCTCCTTCAATCAACTCTTTAAATCT 539
Qy 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 540 AAGGCAACTATGATGAAGGCTTTGGCACAGACCCACCAAGGATCTATGGGCAAGCAAA 599
Qy 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 600 AATCAAAACGAAGAGATTTTGGAGAGACCCAGCTTCGCAATGCAAGGGAGACCCCT 659
Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAsnMet 508
Db 660 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGTGCGTGTCTCTGCTGCTGCAAGTATG 719
Qy 509 GluAlaLysAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeu 528
Db 720 GAACCCAAAGCCCTCTCTCAGAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAAAGCTG 779
Qy 529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGly 548
Db 780 AGGATCGCTGGCCACCCCTCACTGAACTTGGAAAGTTCAAGAAAGTGCCTTTGGAGGAGGG 839
Qy 549 IleLysMetSerLysProLysTyrProProGluAspGluLysSerLysProGluValPro 568
Db 840 ATCAAAATGTCAAAAGCCCAAAATGGCTCTCTGAAAGACGAAATCAGCAAGCCCGAAGTTCT 899
Qy 569 GluAspValAlaSerLeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSer 588
Db 900 GAGGATGTCTGATCTAGATCTGAAGAGAGCTTAAGACGATCTTCTTCTACTCAAGGAAGAAGC 959

Db 1118 AAGGGCAACTATGATGAAGCTTTGGGCACAGACACACAGGATCTATGGGCAAGCAGA 1177
Qy 469 AsnGluAsnGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1178 AATGAAACCAAGAGATTTTGGAGAGACAGCCAGCTTCAAAATGCAAGGGAGACCCCT 1237
Qy 489 HisSerProGlyValGluAspAlaProIleAlaIysValGlyValLeuAlaAlaSerMet 508
Db 1238 CACAGCCAGGGTGAAGATGCCCTATTGCTAAGTGGTGTCTCTGGCTGCAAGTATG 1297
Qy 509 GluAlaLysAlaSerSerGlnGlnGluLysGluLysProAlaGluThrLysLysLeu 528
Db 1298 GAAGCCAAAGCCTCTCTACAGAGGAGAGAGACAGCCAGCTGAACCAAGAGAGCTG 1357
Qy 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1358 AGGATCCCTGGCCACCCCTGAACTTGAAGTTCAGAAAGTCCCTTGGAGGAGGG 1417
Qy 549 IleIysMetSerLysProLysTrpProProGluAspGluIleSerLysProGluValPro 568
Db 1418 ATCAAAATGTCAAGCCCAATGGCTCTGGAAGAGCAATATCAGCAAGCCGGAAGTTCCT 1477
Qy 569 GluAspValAspLeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSer 588
Db 1478 GAGGATGTCATCTAGATCTGAAGAGCTTAAGACGATCTTCTACTGAAGGAAGAGC 1537
Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1538 CGCCCATCTACTAGCAGCTTCTATTCAAGCACCTCTGTCAAGAGCCCAAACTGTG 1597
Qy 609 SerProPheIleArgLysGlyTrpSerMetSerGluGlnSerGluSerValGlyGly 628
Db 1598 TCCCCACCTATCAGGAAAGCTCGAGCATGTCAAGACAGATGAAGTCTGTGGTGA 1657
Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1658 AGASTTGCAGAAAGGAAACAGTGGAAATGCCAAGCTTCTAAGAAAGATGGGAATGTG 1717
Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1718 GGAAAAACACCTGGCAAAACAAAGAAATCTAAAGGAGACAGAGGAGAGAAAGTAAGGA 1777
Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1778 GGTCTAGTTTGGAGATGAGAAATGAGATCTTTAGAAAATGGTGCAGACTCCGATGAA 1837
Qy 689 AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluPheThrGlnAsnGlnLysSerGlnAsp 708
Db 1838 GATGATTAACAGCTTCTCTCAAAACAAATCTCCACAGAAACCCCAAGTCTCTGAATTGGTGG 1897
Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrGlnAsnGlnLysSerGlnAsp 728
Db 1898 AGTTTGTAGACACACCTTTGCTGAGAAATTCATCTACTCAGATCAGAAATCCCAAGAT 1957
Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1958 GTGGAATCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATTAAGAGA 2017
Qy 749 AsnArgTyTyAspGluAspGluAspGluGlu 759
Db 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAG 2050
RESULT 12
ID AA233566 standard; cDNA; 2783 BP.
XX
AC AA233566;
DT 08-DEC-1999 (first entry)
XX Human breast tumour-associated EST 26.
DE Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW

KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX Homo sapiens.
XX DE19813835-A1.
XX 23-SEP-1999.
XX 20-MAR-1998; 98DE-1013835.
XX 20-MAR-1998; 98DE-1013835.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX WPI: 1999-528979/45.
XX P-PSDB; AAY48487.
XX Human nucleic acid sequences and protein products from normal breast
XX tissue, useful for breast cancer therapy
XX Claim la; 113-114; 206pp; German.
XX This invention describes novel human nucleic acid sequences from normal
XX breast tissue which have cytostatic activity. The nucleic acid sequences
XX can be used to produce and isolate full-length gene sequences. They can
XX be used to express proteins, which can be used as tools to find an
XX activity against breast cancer. The sequences can be used in sense or
XX antisense form. They are especially useful for medicaments for gene
XX therapy to treat breast cancer and for treating illnesses associated
XX with fat metabolism. AA233541-233610 represent expressed sequence tags
XX described in the method of the invention.
SQ Sequence 2783 BP; 901 A; 527 C; 608 G; 747 T; 0 other.
Alignment Scores:
Pred. No.: 7,26e-167 Length: 2783
Score: 2455.00 Matches: 469
Percent Similarity: 99.79% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 62.52% Indels: 0
DB: 20 Gaps: 0
US-09-890-549-4 (1-759) x AA233566 (1-2783)
Qy 289 AsnGluLeuLysAlaSerGlyGlyLeuLysIleHisLysMetGluGlnLysGluAsn 308
Db 40 AATGAGCTGAAGCAGTGGTGGCAATCAAAATTCATAAATGGAGCAAAAGAGAAAT 99
Qy 309 ValProProGlyProGluValCysIleThrHisGlnGluGlyGlyLysIleSerAlaAsn 328
Db 100 GTGCCCCCAGGTCTGAGGTCTGCATCACCCTCAGCAAGGGGAAAGATTTCTGCAAT 159
Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
Db 160 GAGAATAGCTGGCAGTCCGTTCACCCCTGCCAAGATGACTCCCGTACTCCAGGTT 219
Qy 349 LysSerGluValGlnGlnProValHisProLysProLysSerProAspSerArgAlaSer 368
Db 220 AAGAGTGGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 279
Qy 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 280 AGTCTTCTGAAGTCTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 339
Qy 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 340 ACCTGGTGGAAATGTCAGAGACAGTCTATCCAAATGAGCGTCTCTTGGCCCAACAGCAG 399
Qy 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThr 428
Db 400 GTGTTTCAATCAGCTGCTTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACA 459

QY 429 TyrAlaSerLeuHisGlyArgGileTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 460 TATGCATCTTTACATGGAAGAAATCTATTGTAAGGCTCACTCAATCAACTCTTTAAATCT 519
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 520 AAGGGCACTATGATGAAGCTTTGGGCACAGACCACACAGGATCTATGGCAAGCAAA 579
QY 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 580 AATGAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTCAAATGCAAGGAGAGCCCT 639
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 640 CACAGCCCGGGGTAGAGATGCCCTTATTGCTAAGGGGGTGTCTGCTGGCTGAAGTATG 699
QY 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 700 GAAGCCAAAGCCCTCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 759
QY 529 ArgIleAlaTrpProProThrProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 760 AGGATCGCTGGCCACCCCTCACTGAATTTGGAAGTTTCAGGAAGTGCCTTGGAGAGGG 819
QY 549 IleLysMetSerLysProLysTyrProProGluAspGluLysSerLysProGluValPro 568
Db 820 ATCAAAATGTCAAGCCCAATGAGCTCTCTGAAGAGCAAAATCAGAGCCGAGTTCCT 879
QY 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerSerLysLysGluArgSer 588
Db 880 GAGATGTCATCTAGATCTGAGAGAGCTAAGAGATCTTCTTCACTGAAGGAAAGAGC 939
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 940 CGCCATTCAGTGTAGCAGCTTCAATTTCAAGCACCTCTGTCAAGAGCCCAAAATCTGTG 999
QY 609 SerProIleArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1000 TCCCCACCTATCAGGAAGGCTGGAGCATGTCAAGAGAGAGTCAAGAGTCTGTGGGTGA 1059
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1060 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTAAGAGAGATGGGAATGTG 1119
QY 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1120 GGAATAACACCTGGCAAAACAAAGATCTTAAGGAGAGAGAGAGAGAGAGAGAGAGAA 1179
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1180 GGTCAATAGTTGGAGATGGAGATGAGAACTCTGTAGAAAATGGTCCAGACTCCGATCAA 1239
QY 689 AspAspAsnSerPheLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1240 GATGATAACAGCTTCTCTCAAAACAAATCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 1299
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 1300 AGTTTTGTAGACACACCTTTGCTGAGAGATTCCTACTCTCAGATCAGAAATCCAGGAT 1359
QY 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnLysArg 748
Db 1360 GTGAACTCTGGAGGAGAGAGAGTGTCAAAGAGCTCTCTGTGGAGAGAGAGAGAGAG 1419
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1420 AATCGGTATTATGATGAGGATGAGGATGAAGAG 1452

RESULT 13

AAS18588

ID AAS18588 standard; cDNA; 1567 BP.

XX

AC AAS18588;
XX 12-MAR-2002 (first entry)
XX cDNA encoding human sterol regulatory element binding protein 3.
DE XX Human; sterol regulatory element binding protein 3; hSREBP-3;
KW hypothalamus; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
CDS 169..154
/*tag= a
/product= "Sterol regulatory element binding protein 3,
hSREBP-3"
XX CN1309182-A.
XX 22-AUG-2001.
XX 17-FEB-2000; 2000CN-0111698.
XX 17-FEB-2000; 2000CN-0111698.
XX (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.
XX Li Y, Xu S, Ren S;
XX WPI; 2002-011822/02.
XX P-PSDB; AAU10979.
XX Cholesterol regulatory factor binding protein and its coding sequence -
XX Claim 1; Page 22; 27pp; Chinese.
XX The invention relates to a novel human sterol regulatory element binding
protein 3 (hSREBP-3) expressed in human normal hypothalamic tissue and
its coding sequence. Also described is the process for preparing the
protein and nucleic acid sequence, and the method for detecting hSREBP-3
nucleic acid sequence and polypeptides. The present sequence represents
the coding sequence of human hSREBP-3 as described in the invention.
XX Sequence 1567 BP; 513 A; 326 C; 391 G; 337 T; 0 other;

Alignment Scores:

Pred. No.: 9.47e-167 Length: 1567
Score: 2449.00 Matches: 468
Percent Similarity: 99.79% Conservative: 2
Best Local Similarity: 99.36% Mismatches: 1
Query Match: 62.36% Indels: 0
Gaps: 24

US-09-890-549-4 (1-759) x AAS18588 (1-1567)

QY 289 AsnGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAsn 308
Db 127 AATGAGCTGAAGCCAGTGTGGCAAAATCAAAATTCATATAATGGAGCAAAAGAGAAAT 186
QY 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328
Db 187 GTGCCCCCAGGTCTCTGAGGTCTGCATCCCATCAGGAGGGGAAAGATTTCTGCAAAAT 246
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
Db 247 GAGAATAGCTGGCAGTCCGTTCACCCCTCCGAGAGATGACTCCCGTACTCCAGGTT 306
QY 349 LysSerGluValGlnGlnProValHisProLysProLysSerProAspSerArgAlaSer 368
Db 307 AAGAGTGAAGTTCACAGAGCTCTCCATCCCAAGCCACTAAGTCCAGATTCACAGAGCTCC 366
QY 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 367 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 426
Qy 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 427 ACCTCGGTGGAAATGTCAGAACAGAGTCTATCCATGGAGGCTCTCTTGGCCCAACACGAG 486
Qy 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428
Db 487 GTGTTTCACATCAGCTGCTTCGTTGCTCTTATTCGCAACAACAATCAGTCTAGGAACA 546
Qy 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 547 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACCTCAATCAACTCTTTAATCT 606
Qy 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468
Db 607 AAGGCAACTATGATGAAGGCTTTGGCCACAGACACACAGGATCTATGGCAAGCAAA 666
Qy 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 667 AATGAAACAGAGAGATTTGGAGAGACACGCCAGCTTGCATGCAAGGGAGAGCCCT 726
Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 727 CACAGCAAGGGGTAGAAGATGCCCTATTGCTAAGGTGGTGTCTTGGCTGCAAGTATG 786
Qy 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 787 GAAGCCCAAGCCCTCTCTCAGCAAGAGAGAGCAACAGCCGAGCTGAACCAAGAGCTG 846
Qy 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGly 548
Db 847 AGATGCCCTGGCCACCCACCTGAACTTGGAGTTTCAGGAAGTGCCTTGGAGGAGGG 906
Qy 549 IleLysMetSerLysProLysTyrProProGluAspGluIleSerLysProGluValPro 568
Db 907 ATCAAAATGTCAAGGCCCAATGGCTCTCTCAGAGCAAGATCAGCAAGCCCGAAGTTCT 966
Qy 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerLeuLysGluArgSer 588
Db 967 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAAAGAGC 1026
Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1027 CGCCATCTACTGTAGCAGCTTCAATTCAAAGCACCTCTGTCAAGAGCCCAAACTGTG 1086
Qy 609 SerProPheIleArgLysGlyTyrSerMetSerGluGlnSerGluSerValGlyGly 628
Db 1087 TCCCCACCTATCAGGAAGGCTGGAGCATCTCAGACGAGTGAAGAGTCTGTGGGTGA 1146
Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1147 AGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAGCTTCTAAGAGAGATGGGAATGTG 1206
Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1207 GGAAGAAACACCTGGCAAAACAAGAAATCTAAAGGAGACAGAGGAGAGAGTAAGAA 1266
Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1267 GGTCAATAGTTGGAGATGAGAAATGAGATCTGTAGAAAATGGTGCAGACTCCGATGAA 1326
Qy 689 AspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1327 GATGATAACAGCTTCTCAACACACATCTCCACAGAACCCCAAGTCTCTGAATTTGGTGG 1386
Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 1387 AGTTTTGTAGACAACACCTTTGTGAAGAATTCCTACTACTCAGATCAGAAATCCAGAGAT 1446
Qy 729 ValGluLeuTrpGluGlyGluValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1447 GTGGAACTCTGGAGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 1506

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1507 AATCGTATTATGATGAGGATGAGGATGAAGAG 1539

RESULT 14

AAAC59489
ID AAAC59489 standard; cDNA; 2267 BP.
XX AAAC59489;
XX XX
DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein gene 41 SEQ ID NO:51.

XX Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnary; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative; ss.

XX Homo sapiens.

OS WO200056755-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06830.

XX 19-MAR-1999; 98US-0125361.

XX 10-DEC-1999; 98US-0169910.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-587661/55.

XX P-PSDB; AAB34132.

XX New isolated nucleic acid molecules encoding 49 human secreted proteins
PT used for preventing, treating or ameliorating medical conditions, for
PT diagnosing pathological conditions or as food additives or
PT preservatives

XX Claim 1; Page 357-358; 419pp; English.

XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
CC AAB34216 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissue
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used

CC in the exemplification of the present invention.
XX
SQ Sequence 2267 BP; 741 A; 416 C; 490 G; 617 T; 3 other;

Alignment Scores:
Pred. No.: 1,586-104 Length: 2267
Score: 1586.00 Matches: 306
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.39% Indels: 0
DB: 21 Gaps: 0

US-09-890-549-4 (1-759) x AAC59489 (1-2267)

QY	454	GlulGlyPheGlyHisArgProHisLysAspLeuTTPAlaSerLysAsnGluGlu	473
DB	32	GAAGGTTTGGCCACAGACACACAGGATCTATGGCAAGCAAAATGAAACGAGAG	91
QY	474	IleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyVal	493
DB	92	ATTTTGGAGACACAGCCAGCTTGCATGTAAGGAGAGCCCTCACAGCCCGGGTA	151
QY	494	GlulAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSer	513
DB	152	GAAGATGCCCTATTGCTAAGTGGTGTCTCTGGCTGCAAGTATGGAAGCCCAAGCCCTCC	211
QY	514	SerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTTPPro	533
DB	212	TCTCAGCAGGAGAGGAGACAGCAGCTGAACCAAGAGCTGAGGATCGCTGGCCA	271
QY	534	ProProThrGluLeuGlySerSerGlySerAlaLeuGluGlyIleLysMetSerLys	553
DB	272	CCCCCCTCACTGAAGTTCAGGAAGTTCAGGAAGGAGGATCAAAATGTCAAG	331
QY	554	ProLysTTPProProGluAspGluLysSerLysProGluValProGluAspValAspLeu	573
DB	332	CCCAATGGCTCTCTGAGACGAATCAGCAGCCGAGTTCCTGAGATGTGATCTA	391
QY	574	AspLeuLysLysLeuArgSerSerSerLeuLysGluArgSerArgProPheThrVal	593
DB	392	GATCTGAAGAGAGCTTAAGACGATCTTCTCACTGAAGGAAAGACCCCTTCACTGTA	451
QY	594	AlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProIleArg	613
DB	452	GCAGCTTCAATTCAGACCTCTGTCAAGAGCCCAAAATGTGTCCACCTATCAGG	511
QY	614	LysGlyTTPSerMetSerGluGlnSerGluGlnSerValGlyArgValAlaGluArg	633
DB	512	AAAGCTGGAGCATGTCAGACAGAGTGAAGAGTCTGTGGTGGAGAGTTCAGAAAGG	571
QY	634	LysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTTP	653
DB	572	AAACAAGTGGAAATGCCAAGCTTCTAAGAAGATGGGATGTGGGAAACAAACCTGG	631
QY	654	GlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGlu	673
DB	632	CAAAACAAGAAATCTAAAGGAGACAGCAGGAGAGAGTGAAGAGTATAGTTTGGAG	691
QY	674	MetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPhe	693
DB	692	ATGAGAGATGAGATCTTTGTAAGAAATGGTGCACTCCGATGAAGATGAATACAGCTTC	751
QY	694	LeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTTPSerSerPheValAspAsn	713
DB	752	CTCAACAACAATCTCCACAAGACCCCAAGTCTCTGAATTTGTCGATTTTGTAGACAAC	811
QY	714	ThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTTPGlu	733
DB	812	ACCTTTCTGAAGAATCTCACTACTCAAAATCAGAAATCCCAAGATGTGGAACTCTGGAG	871
QY	734	GlyGluValValLysGluLeuSerValGluGluGlnLysArgAsnArgTyrTyrAsp	753
DB	872	GGAGAGTGTCAAAGAGCTCTCTGTGGAGAGACAGATAAAGAGAAATCGTTATTATGAT	931

QY 754 GlulAspGluAspGluGlu 759
DB 932 GAGGATGAGGATGAGAG 949

RESULT 15
AAC93483
ID AAC93483 standard; cDNA; 2158 BP.
XX
AC AAC93483;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human secreted protein gene 5 SEQ ID NO:15.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX ophthalmological; vulnary; autoimmune disease; rheumatoid arthritis;
XX hyperproliferative disorders; cancer; cardiovascular disorder;
XX cardiac arrest; cerebrovascular disorder; nervous system disorder;
XX Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.

XX Homo sapiens.

XX WO2000061626-A1.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US09066.

XX 09-APR-1999; 99US-0128698.

XX 20-JAN-2000; 2000US-0176926.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-619227/59.

XX P-PSDB; AAB51831.

XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing or ameliorating medical conditions and used for
XX food additives or preservatives -

XX Claim 1; Page 435; 516pp; English.

XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
XX human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
XX AAB51927 represent alternative polypeptides encoded by the genes, and
XX amino acid sequences with which they share homology. The genes and
XX proteins have activities dependent on the tissues and cells in which they
XX are expressed. Examples of their activities include immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
XX virucide; fungicide; ophthalmological; and vulnary. The secreted
XX proteins, polynucleotides, antagonists and agonists may be useful in
XX treating, preventing and/or diagnosing diseases and disorders such as
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angioneurosis, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi; and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,
XX carbohydrate, vitamins, minerals, cofactors and other nutritional
XX components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
XX used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX Sequence 2158 BP; 714 A; 383 C; 458 G; 603 T; 0 other;

Alignment Scores:

Pred. No.: 5,62e-92 Length: 2158
Score: 1411.00 Matches: 286
Percent Similarity: 98.29% Conservative: 1
Best Local Similarity: 97.95% Mismatches: 0
Query Match: 35.93% Indels: 5
DB: 21 Gaps: 0

US-09-890-549-4 (1-759) x AAC93483 (1-2158)

QY 473 GluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGly 492
DB 1 GAGATTTTGAGAGAGACAGCCCGACCTTGCATAATGCAAGGGAGACCCCTCACAGCCAGGG 60
QY 493 ValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAla 512
DB 61 GTAGAAGATGCCCTATTGCTAAGGTGGTGTCTCGCTCAAGTATGGAAGCCCAAGGCC 120
QY 513 SerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrp 532
DB 121 TCCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCCTGG 180
QY 533 ProProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSer 552
DB 181 CCACCCCCCACTGAATTTGAAGTTGGAAGTTCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCA 240
QY 553 LysProLysTrpProGluAspGluIleSerLysProGluValProGluAspValAsp 572
DB 241 AAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCGAAGTTCCTGAGGATGTGGAT 300
QY 573 LeuAspLeuLysLysLeuArgSerSerSerLeuLysGluArgSerArgProPheThr 592
DB 301 CTAGATCTGAAGAAGTAAGACGATCTTCTCAGTGAAGAAAGAGCCGCCCATTCAC 360
QY 593 ValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPhe 612
DB 361 GTAGCAGCTTCAATTCACAGCACCTCTGTCAAGAGCCCAAAACCTGTGTCCCACTATC 420
QY 613 ArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGlyArgValAlaGlu 632
DB 421 AGGAAGGCTGGAGCATGTCAAGCAGCATGAGAGTCTGTGGGTGGAAGAGTTGCAGAA 480
QY 633 ArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThr 652
DB 481 AGCAAAACAAGTGGAAATCCCAAGSCTTCTAAGAGAAATGGAAATCTGGGAAACAAACC 540
QY 653 TrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeu 672
DB 541 TGGCAAAACAAAGAATCTAAAGGAGAGACAGGGAAGAGTAAGGAAGGTCAAGTTTG 600
QY 673 GluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSer 692
DB 601 GAGATGGAGAAATAGAAATCTTGTAATAATGGTGCAGACTCCGATGAAGATCATACAGC 660
QY 693 PheLeuLysGlnGlnSerPro--GlnGluProLysSerLeuAsnTrpSerSerPheVal- 711
DB 661 TTCCTCAAAACAACATCTCCACCAAGAACCCCAAGTCTCTGAATTTGGTGGAGTTTGT 720
QY 712 AspAsnThrPhe-AlaGluGlu-PheThrThrGlnAsnGlnLysSerGlnAspValGluL 731
DB 721 GACAACACCTTTGGCTGAAGAATTTCACTACTCAGAATCAGAAATCCAGGATGTGGAAC 780
QY 731 euTrpGluGlyGluValLysGluLeuSerValGluGluGlnIleLysArgAsnArg 751
DB 781 TCTGGAGGGAGAGTGTGTAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGGT 840
QY 751 yrTyAspGluAspGluAspGluGlu 759
DB 841 ATTATGATGAGGATGAGGATGAGAG 866

Search completed: January 6, 2004, 22:38:06
Job time : 537 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:07:02 ; Search time 3923 Seconds
(without alignments)
4702.293 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSPFNRQRTSLSLRVT.....LSVEQIKRNRYDEDEEE 759

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spool/US0980549/runat_06012004_094752_19831/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPECI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=EXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980549@cgn_1_1.2810/runat_06012004_094752_19831 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estro:*
7: em_hic:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_eston:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2955	75.2	3242	11	AK049350 Mus muscu
2	2718	69.2	3108	11	AK031698 Mus muscu
3	2713	69.1	3100	11	AK028186 Mus muscu
4	2707	68.9	3160	11	AK085065 Mus muscu
5	1532.5	39.0	1075	12	BM909158 AGENCOURT
6	1401	35.7	1142	12	EM914155 AGENCOURT
7	1294	33.0	971	10	BE410108 AGENCOURT
8	1252.5	31.9	989	12	BM52304 AGENCOURT
9	1243	31.7	768	10	BG476496 AGENCOURT
10	1222	31.1	918	13	BU850203 AGENCOURT
11	1219.5	31.1	787	13	BU861039 AGENCOURT
12	1212	30.9	847	13	BU803212 AGENCOURT
13	1209	30.8	828	12	BI457843 AGENCOURT
14	1199.5	30.5	770	12	BI089829 AGENCOURT
15	1193	30.4	931	12	BM449793 AGENCOURT
16	1190	30.3	1120	10	BG120125 AGENCOURT
17	1186.5	30.2	993	12	BM458935 AGENCOURT
18	1186	30.2	884	14	CD171644 AGENCOURT
19	1181.5	30.1	951	10	BF528886 AGENCOURT
20	1172	29.8	715	9	AL048161 DKFZP586K
21	1170	29.8	765	12	BI557269 AGENCOURT
22	1148	29.2	941	10	BG112799 AGENCOURT
23	1126	28.7	643	13	BQ632628 AGENCOURT
24	1120.5	28.5	936	13	BQ713158 AGENCOURT
25	1114	28.4	884	13	BU527038 AGENCOURT
26	1106	28.2	634	10	BE729955 AGENCOURT
27	1103	28.1	777	28	AQ314576 AGENCOURT
28	1101	28.0	925	13	BU856577 AGENCOURT
29	1069	27.2	637	14	CA328969 AGENCOURT
30	1050.5	26.8	752	10	BE897604 AGENCOURT
31	1043	26.6	1095	10	BG112232 AGENCOURT
32	998	25.4	717	10	BF752641 AGENCOURT
33	991	25.2	765	14	CD512151 AGENCOURT
34	975	24.8	846	12	BG911181 AGENCOURT
35	969	24.7	747	14	CB963617 AGENCOURT
36	967.5	24.6	767	14	CB988778 AGENCOURT
37	959.5	24.4	931	13	BQ212132 AGENCOURT
38	955	24.3	728	13	BU631767 AGENCOURT
39	952.5	24.3	809	10	BE621212 AGENCOURT
40	945	24.1	841	10	BE542098 AGENCOURT
41	939	23.9	576	10	BE005092 AGENCOURT
42	935	23.8	933	10	BG539577 AGENCOURT
43	932	23.7	843	9	AA909018 AGENCOURT
44	906	23.1	665	12	BI153527 AGENCOURT
45	897.5	22.9	654	14	CB050200 AGENCOURT

ALIGNMENTS

RESULT 1

AK049350

LOCUS

DEFINITION

AK049350

VERSION

KEYWORDS

SOURCE

ORGANISM

AK049350

GI:26340071

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AK049350 3242 bp mRNA linear HTC 05-DEC-2002
Mus musculus ES cells cDNA, RIKEN full-length enriched library, full
clone: C330026L07 product: epithelial protein lost in neoplasm, full
insert sequence.

AK049350

AK049350.1 GI:26340071

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carninci, P. and Hayashizaki, Y.	
AUTHORS		High-efficiency full-length cDNA cloning	
TITLE		Meth. Enzymol. 303, 19-44 (1999)	
JOURNAL		99279253	
MEDLINE		PUBMED	
REFERENCE	2	10349636	
AUTHORS		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL		Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE		20499374	
PUBMED		11042159	
REFERENCE	3	11042159	
AUTHORS		Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL		Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE		20530313	
PUBMED		11076861	
REFERENCE	4	11076861	
AUTHORS		Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Kachiwa, H., Quackenbush, J., Schiraldi, L. M., Staehli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, P., Bojunga, N., Carninci, P., De Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.	
TITLE		Functional annotation of a full-length mouse cDNA collection	
JOURNAL		Nature 409 (6821), 685-690 (2001)	
MEDLINE		21085660	
PUBMED		11217851	
REFERENCE	5	11217851	
AUTHORS		The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
TITLE		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL		Nature 420, 563-573 (2002)	
REFERENCE	6	(bases 1 to 3242)	
AUTHORS		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome	

COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in RIKEN contributed to prepare mouse tissues.

COMMENT

Please visit our web site for further details.

COMMENT

URL: http://genome.gsc.riken.go.jp/

COMMENT

URL: http://fantom.gsc.riken.go.jp/

FEATURES

Source

1. .3242
/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C330026L07"
/db_xref="taxon:10090"
/clone="C330026L07"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
105. .2366
/note="unnamed protein product; epithelial protein lost in neoplasm (MGI:MGI:1920992, GB|NM_023063, evidence: BLASTN, 99%, match=3086)
putative"
/codon_start=1
/protein_id="BAC33699.1"
/db_xref="GI:26340072"
/translation="MESTFNRNROWTSLRVTAKELSLVKNKSSAIVEIFSKYKAA
AEEANMERKKNPESLPQFRFRGLSVLKKWNPVAGAFPTDSLPNSSEGGHTAD
YPPAEVTDKPAQGVADREHTQKPRFGSRPEAVIOSRYPSNSHDFKAQATESQK
MENCGLSRHEAEKPTGENTETSGKIEKYNVPLNRLKMPKEGHEHQTKSLWTQSRN
AVSQSSPASTYNELKTSKTHKWEKENVPPGPEAYSVHQEGSKYSTTNSVALS
VPAEDTCSQVSEAAQPMHPKLPSPDARTSLSPSSPKTKAKFOAPAKESCVCEQ
KTYVMERLANQVFIISCFRCYCNKLSLGTYSALHGRVYCKPHNQLFKSKGNV
AGGRGHQKHDLWASKDNEETLGRPAQPPNAGESPHSPGVEDAPIAKGVLAASMEA
KASSORREDKPAETKRLIAMPPEALGGSGSALEGGIKVSKPKPPEDDVCKTEAP
EDVDLDLKKLRSSSLKERSRPFVAAASFTSSIKSPASSPSLRKMGWSESEEEFG
GGIATMERKQTNARPSGKENVKSWQGEVPRSDRSFELESNFMENGANIAE
DDNHVHAQSPLEPEAPGWSGFDVTTAAKEFTTQKQSDQVDFWEGEVRELSVEEQI
KRNRYDEDEDEE"

CDS

621 t
961 a 797 c 863 g 621 t
ORIGIN
BASE COUNT
Alignment Scores:
Pred. No.: 1.62e-218 Length: 3242
Score: 2955.00 Matches: 575
Percent Similarity: 83.55% Conservative: 60
Best Local Similarity: 75.66% Mismatches: 117
Query Match: 75.25% Indels: 8
DB: 11 Gaps: 3
US-09-890-549-4 (1-759) x AK049350 (1-3242)

QY	1	MetClusSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuAtrgValThra	20
Db	105	ATGAATTCGATCCATTTAATAGACGCCAGTCGACTCCCTGTCATTGAGAGTAACAGCC	164
QY	21	LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys	40
Db	165	AAAGAGCTTTCTTTGTCACACAGAACAGTCACTCCGCAATTTGTAATAATTTCTCAAG	224
QY	41	TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu	60
Db	225	TACCAGAGGCTCTCTGAGAGGCCACATCGAAAGAGAGAAAATTAACCTGAAAGCTCG	284
QY	61	SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly	80
Db	285	CCCCAGCACTTTAGAGAGGGACCCCTGCTCTGTGTTAAAGAGAGTGGGAGAACCCGGTG	344

QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluLeuLeuArgHisArg 100
DB 345 GCTGGGGCAGATTCCACACAGACTCACTGCCAAACAGCAGCAGTGGGGTGGGCACACA 404
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
DB 405 GCGGACTACCTCTCTCAAGTACGGACAAAGCCTGCTCTCTGGAGTCAAGACCTGACCGG 464
QY 121 GluGluGlnHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB 465 GAAGAGCACACCCAGCCCAACCTAGATTGGATCTCGTCCGGAAGACAGTTATCCAGAGC 524
QY 141 ArgTyrProHisLeuLysAspGlyGluAspLeuLysAspHisSerThrCluSerLysLys 160
DB 525 CGGTATCTCGCTCAGAGAACACCCAGCATTTAAAGCCCGCCAGCAGGAGCCCAAAA 584
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLeuSerGluAsn 180
DB 585 ATGGAATACTGTCTGGAGATTCCAGGCATGAAGCAGAGAACCCAGAGACGGCGCAAAAC 644
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB 645 ACAGAAACTTCAGGCAAAATAGAGAAATACAAACGTTCCACTGAATAGACTGAAGATGATG 704
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 705 TTGTGAAGAAGGTGAACACAAACCAACAGAGTCTCTGACCCCAAGCCGGAATGCGGGT 764
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluLysIleGlyProGln 240
DB 765 GGAAGAGGCTCTGTGAATACTGTCTCTGGATGACTGGGAATAGTTCGGGTTCAT 824
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 825 TTGTCTCTCTGTCATCAACTCGGAGAAACAGAGAGTAAGAGAACTCTCGAGCTGCCA 884
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
DB 885 CCCCTCTCAGAACTCTCATAAAGACCCGATGGCCAAAGTACCAGCTGCAGTGTCAAAG 944
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysAlaSerGlyGlyLysIleLysIle 300
DB 945 CAGAGACCCAGCCAGCTACAAATAGCTGTAACCACTG-----GAAAGCAAACT 998
QY 301 HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
DB 999 CATAAATGGGAACAGAGAGAAATGTCGCCCGAGGTCCCGAGGCTACAGGCTCCATCAG 1058
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1059 GAAGGAAGCAAGGTCTTACAACTGAGAATAGTCTAGTGGCCCTCTCTGCTCCCTGCTGAA 1118
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1119 GATGACACCTGTAATCTCCAGGTGAAGAGCAGGCCAGCAGCCTATGCACCTTAAGCCG 1178
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
DB 1179 CTGAGTCTCGAGCCAGGACCTCCAGTCTTCGGAAAGTCTCTCTCCAAACAGCGAAG 1238
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1239 AAGTTTCAGGCGCGGCAAGAGAGAGTGGTGGAGTGTGAGAAGACGGGTGACCCCATG 1298
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1299 GAACGGTCTCTGGCCAAACAGCAGGTGTTTCATCAGCTGTTTCCGATCTCTACTGC 1358
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1359 AACACAGCTCAGCTAGAGAACATATGCATCTTATATGAGCAATCTACTCTAAGCCT 1418

QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1419 CACTTCAATCAACTCTTCAAGTCCAAAGGCAACTATACAGAGGCTTTGGGCACAAGCAG 1478
QY 461 HisLysAspLeuTyrAlaSerLysAsnGluGluLeuLeuGluArgProAlaGln 480
DB 1479 CATAGAGATCTGTGGGCAAGCAAGAGCGCAATAGGAGACTTTTGGGAGACCGAGCCAG 1538
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1539 CCGCTAATGCAGGGAGAGCCCCCATAGCCGGGTGTAGAGATGCCCATCGCCCAAG 1598
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
DB 1599 GTCCGCGTGTGGCGCAAGTATGGAAGCAAGCCCTCTTCTCAGAGGAGAGGAGGAT 1658
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySer 540
DB 1659 AAGCCTCGGAGACCAAGAGCTAGGATCGCTGGCTCCGCCAGCGGAGCTGGCGGT 1718
QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProProGluAsp 560
DB 1719 TCCGAAAGTCCCTGGAGGAGGAGATCAAGTATCGAAGCCCAAGTGGCTCCGAGGAT 1778
QY 561 GluLeuSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
DB 1779 GAGCTCTGCAAGCGGAGGCGCCCGAGGATGATAGTCTGACCTGAAGAACTACGCGG 1838
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1839 TCTTCGTCCTGAGGAGAGGAGCGCCCTTTACGCTAGCAGCTTCTTTCGAACCTCT 1898
QY 601 SerValLysSerProLysThrValSerProProLysArgLysGlyTyrSerMetSerGlu 620
DB 1899 TCCATCAAGAGCCCGGAGGCTCTTCCCATCTCTCAGGAAAGGTGGAGCGAGTCCGAG 1958
QY 621 GlnSerGluGluSerValGlyGly---ArgValAlaGluArgLysGlnValGluAsnAla 639
DB 1959 CAGAGTGAGAGGTTTGGAGAGGAGTAGCGACGATGGAAGGAAACAAACGGAAATGCC 2018
QY 640 LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLys 659
DB 2019 AGACCTCTGGGAGAGAGAGATGTGGAAATACCGCTGGCAG----- 2063
QY 660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluLeu 679
DB 2064 GCGAGAGAGTGGCGAGGAGCAAGGACCGTAGTCTCTTTGAGCTGAGAGTGAAGATTTT 2123
QY 680 ValGluAsnGlyAlaAspSerSerAspGluAspAspAsnSerPheLeuLysGlnSerPro 699
DB 2124 ATGGAATAAGGAGCAACATAGCTGAAGATGACACCATGTCCACGACGAGTCCCGG 2183
QY 700 GlnGluProLysSerLeuAsnTyrSerSerPheValAspAsnThrPheAlaGluPhe 719
DB 2184 TTAGAACCAGGCTCCGGGTGGTCTGGCTTTGTAGACACCGCTGCTTAAGAAATTC 2243
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGlyValValLysGlu 739
DB 2244 ACTACCCAGAAATCAAAATCCCAAGATGCGGTTCTGGAGGAGAGAGTGTCCGAGAG 2303
QY 740 LeuSerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluGlu 759
DB 2304 CTGTCTGTGGAGAAACAGATAAAGAGAAACCGGTACTATGATGAGGACGAGATGAAGAA 2363

RESULT 2

AK031698

LOCUS

DEFINITION

Mus musculus 13 days embryo male testis cDNA, RIKEN full-length

enriched library, clone:6030492D07 product:epithelial protein lost

in neoplasm, full insert sequence.

ACCESSION

AK031698

VERSION

AK031698.1

GT:26327552

KEYWORDS

HTC; CAP trapper.

AK031698 3108 bp mRNA linear

Mus musculus 13 days embryo male testis cDNA, RIKEN full-length

enriched library, clone:6030492D07 product:epithelial protein lost

in neoplasm, full insert sequence.

ACCESSION

AK031698

VERSION

AK031698.1

GT:26327552

KEYWORDS

HTC; CAP trapper.

SOURCE	Mus musculus (house mouse)	TITLE	Muramatsu, M. and Hayashizaki, Y.
ORGANISM	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	JOURNAL	Direct Submission
REFERENCE	1 Carninci, P. and Hayashizaki, Y.		Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	FEATURES	Location/Qualifiers
MEDLINE	20499374		1. 3108
PUBMED	11042159		/organism="Mus musculus"
REFERENCE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Iashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		/mol_type="mRNA"
	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		/strain="C57BL/6J"
TITLE	Genome Res. 10 (11), 1757-1771 (2000)		/db_xref="FANTOM_DB:6030492D07"
JOURNAL	20530913		/db_xref="taxon:10090"
MEDLINE	11076861		/clones="6030492D07"
PUBMED			/sex="male"
REFERENCE	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischer, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schram, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmshurst, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.		/tissue type="testis"
	Functional annotation of a full-length mouse cDNA collection		/clone_lib="RIKEN full-length enriched mouse cDNA library"
TITLE	Nature 409 (6821), 685-690 (2001)		/dev_stage="13 days embryo"
JOURNAL	21085660		451. .2232
MEDLINE	11217851		/note="unnamed protein product; epithelial protein lost in neoplasm (MGD MG1:1920992, GBI NM_023063, evidence: BLASTN, 99%, match=3086)"
PUBMED			putative"
REFERENCE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		/codon_start=1
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		/protein_id="BAC27520.1"
TITLE	Nature 420, 563-573 (2002)		/db_xref="GI:26327553"
JOURNAL	6 (bases 1 to 3108)		/translations="MENCLDSDSHEAEKPEPTSENTSTSGKIKYVPLNKLKMFEGK EHNQTKSLWTQSNAGRRUNENNSLDDWEIGAGHLSAFNSKRNLELPLR SETSKIDRWAKYQAQVSKQSSPYNLKTSKTHKWEKENVPPGPAQSVHQG SKVSTTNSLVAISVPAEDDTCSQVSKQAQPMHPLSPDARTSSLPESSSKTK AKFOAPAKESCEQKTVYMERLLANQOVPHISCFCSCYCNKLGLTGYASLGRICY KPKNOLFSGKNGVDGFGHKHDKJWASKDNEETLGRPAQPNAGESPGEVDA PIKVGVLASMEAKASSQREDDKPAETKXLRAPPAELGGSSALEEIKVSKP KWPEDVQTEAPEVDLDLKLRRSSSLKRSRFTVAASFTSISKPKASSPSEL ESENFMENGANIAEDDNHVAQOSPLEPEAPGWSGVDTTAAKEFTTQKXSQDVGFW EGEVWRELSVEEOIKRNRYDEDEEE"
MEDLINE			BASE COUNT 911 a 759 c 827 g 611 t
PUBMED			ORIGIN
REFERENCE	6 (bases 1 to 3108)		Alignment Scores:
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,		Pred. No.: 3,7e-200 Length: 3108
			Score: 2718.00 Matches: 527
			Percent Similarity: 82.70% Conservative: 56
			Best Local Similarity: 74.75% Mismatches: 114
			Query Match: 69.21% Indels: 8
			DB: 11 Gaps: 3
			US-09-890-549-4 (1-759) x AK031698 (1-3108)
			QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
			Db 136 AACCTCTGAAGTGTGCCAGCAGCTTTAGAGAGGAGGACCTGTCTGTGTAAAGAAGAG 195
			QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
			Db 196 TGGAGAACCCCGTGGCTGGGCGAGAAATTCACACAGACTCACTGCCAACAGCAGCACT 255
			QY 96 GluileArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly 115
			Db 256 GAGGTGGGCACACACGCGAGCTACCTCTCTCTGCTGAAGTACCGGACCAAGCCTCTCTCTGGA 315

QY 116 AlaLysAlaAspGlnGluGlnIleHisProArgSerArgLeuArgSerProGlu 135
Db 316 GTGAGAGCTGACCGGAGAGACACACCGCCCAACCTAGATTGATCTGTCGGAA 375
QY 136 AlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuHisSer 155
Db 376 GCAGTTATCCAGAGCCGGTATCTCGCTCAGAGACAGACCCAGCATTTAAAGCCAGGCC 435
QY 156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
Db 436 ACGAGAGCCGAAATAATGGAAACTGTCTGGGAGATTCCAGGCATGAAGCAGAGAGCCA 495
QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195
Db 496 GAGACGAGCGAAACACAGAACTTCAGGCATAATAGAGAAATACACGTTCCACTGAAT 555
QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
Db 556 AGACTGAAGATGATTGTTCAGAAAGGTGAACACCAACCAAGAGTCTCTGGACCCAA 615
QY 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGlu 235
Db 616 AGCCGAAATGCGGGTGAAGAGGCTCTCTGAAACAACTGTTCCTGTGATGACTGGAA 675
QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
Db 676 ATAGTTCGGGTCATTGTCTCTCTGATTCACCTCGAGAAACACGAGAGTAAGAGG 735
QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
Db 736 AATCTGGAGTGCACCGCTCTCAGAAACCTCCATAAAGGACCGCATGGCCAAATACAG 795
QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 796 GTGTCAGTCTCAAGCAGAGAGCCAGCCAGCAGCTACACAAATGAGTGAACACCACT 852
QY 296 GlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProGlyProGluVal 315
Db 853 ---GAAAGCAAACTCATTAATGGGAACAGAGAGAGATGTGCCCCAGGTCCCGAGGCC 909
QY 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
Db 910 TGCAGCGTCCATCAGGAAGAGCAAGAGTTCACAACTGAGATAGTCTAGTGGCCCTC 969
QY 336 SerThrProAlaGluAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 970 TCTGTCCCTGCTGAAGATGACACCTGTAACCTCCAGGTGAAGAGGAGCCAGCAGCT 1029
QY 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1030 ATGCACCTTAAGCCGCTGAGTCTCCAGCCAGGACCTCCAGTCTTCCGGAAGTCTCT 1089
QY 376 ProLysAlaMetLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1090 TCCAAAACAGCAAGAGTTTCAGCGCGCGCAAGAGAGAGCTCGGTGGAGTGTGCAAG 1149
QY 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPhe 415
Db 1150 ACGGTGTACCCCATGCAACCGCTCTGCGCAACAGCAGAGGTGTTTCATCATCAGTGTTC 1209
QY 416 ArgCysSerTyrCysAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1210 CGATGCTCTACTGCAACCAAGCTCAGTCTAGGAACATATGCATCCTTACATGGACGA 1269
QY 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1270 ATCTACTGTAGCCCTCACTTCAATCACTCTTCAAGTCCAAAGGCAACTATGACGAGGC 1329
QY 456 PheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluLeuLeu 475
Db 1330 TTGGGCACAGCAGCATTAAGATCTGTGGCAAGCAGAGCCGACATATGAGGACTTTG 1389
QY 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495

Db 1390 GGGAGACAGCCAGCCCTTAATGCGGGAGAGCCCATAGCCCGGTGTAGAAGAT 1449
QY 496 AlaProfileAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1450 GCCCCCATCGCAAGTCTGGCGTCTGGCGCAAGTATGGAAGCCCAAGCCCTCTTCTCAG 1509
QY 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProPro 535
Db 1510 AGGAGAGAGGAGGATAAGCTCGGAGACCAAGAAGCTGAGGATCGCTGCGCTCCGCCA 1569
QY 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysMetSerLysLysPro 555
Db 1570 GCGAGCTGGCGCGTTCGGAAGTCCCTGAGGAAGGGATCAAGATATCGAAGCCCAAG 1629
QY 556 TrpProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1630 TGGCTCCGAGGATGAGCTTCGAGAGGAGGCCCGGAGGATGATGATCTCGACCTG 1689
QY 576 LysLysLeuArgArgSerSerSerLysLysGluArgSerArgProPheThrValAlaAla 595
Db 1690 AAGAAGCTGCGCGCTCTTCGTCTCTGAGAGAGGAGCCCGCTTACGTTAGCAGCT 1749
QY 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProLysLysGly 615
Db 1750 TCCTTTCCAACTCTTCATCAAGAGCCCAAGCCCTCTTCGCCATCTCTCAGCAAGGT 1809
QY 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634
Db 1810 TGGAGCAGTCCGAGCAGAGTGGAGGTTTGGAGGAGGATAGCAGCATGGAAGGAAA 1869
QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1870 CAACGGAATATGCAACCTCTGGGAGAAAGAGAAATGTGGAAATACCGCTGCGAG 1929
QY 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1930 -----GCGAGGAAGTCCCGAGGAGCAGGAGCCAGTCTTCTTTGAGCTG 1974
QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 1975 GAGAGTGAAGATTTATGGAATAATGGAGCAACATAGCTGAAGATGACCAACCATGTCCAC 2034
QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2035 GCACAGCAGTCCCGTTAGAACCCGAGGCTCCGGGTGGTCTGGCTTTGTAGACACACG 2094
QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGluAspValGluLeuTrpGluGly 734
Db 2095 GCTGCTAAAGATTTCACTACCCAGATCAAAATCCCAAGATGTCGGGTTCCTGGAGGGA 2154
QY 735 GluValValLysGluLeuSerValGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2155 GAAAGTGTCTGAGAGCTGTCTGTGGAGAACACATTAAGAGAAACCGGTACTATGATGAG 2214
QY 755 AspGluAspGluGlu 759
Db 2215 GACGAGGATGAAGAA 2229

RESULT 3

AK028186

LOCUS

DEFINITION

AK028186

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AK028186 3100 bp mRNA linear HTC 05-DEC-2002
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610034J22 product:epithelial protein lost
in neoplasm, full insert sequence.

AK028186.1 GI:26389833

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hatada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gliss, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schram, L. M., Staab, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Kono, H., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3100)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tanaka, T., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

COMMENT

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES

source

1. 3100
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2610034J22"
/db_xref="taxon:10090"
/clone="2610034J22"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
448. 2229
/note="unnamed protein product; epithelial protein lost in neoplasm (MGD|GI:1920992, GI|NM_023063, evidence: BLASTN, 99% match=3086)
Putative"
/codon_start=1
/protein_id="BAC25798.1"
/db_xref="GI:26389834"
/translation="MENCGLDSREAEKPTSENTSGTKIEKYNPLRLKMMFEKGEHNTKSLMTQSRNAGRRLLSENNSLDDWEIGAGHSSAFNSKERNLEPRLSETIKRMAYQAQVSKSPASITNELTSKTHKWEKQENPPGPACSVHQEGSKVETSLVALSVPAEDTCSQVKEAQPMHPKPLSPDARTSLGSPSSPKTKAFQAPAKESCVCOTYPMERLNLNQVPHISCFRCYCNKLSGLTYASLHGRIYCKHFNGLFKSGNYDEGFKHOKHDLWASKDNEETLGRPAQPPNAGESPHSPGVEDA PIAKVGLAASMEAKASSQEREDKPAETKRLRIAWPRAELGGSGALSEGKIVSKP KWPEDDYCKTEAPEDVDLKLRLSSSKERSRPTVAASFSTSIKSPKAPSPSL RKWSSEQEEFGGGIATWERKOTENARFSGEKNVCKRWQEEVPRKRSKRSFEL ESEFMENGANIADDDNHVAQSPLESPAPGWSGVDTTAAKEFTTHNQSQDVGFW EGEVYRLSVEEQIKRNYDEDEDE"

CDS

908 a 759 c 824 g 609 t
ORIGIN
Alignment Scores:
Pred. No.: 9.01e-200 Length: 3100
Score: 2713.00 Matches: 526
Percent Similarity: 82.55% Conservative: 56
Best Local Similarity: 74.61% Mismatches: 115
Query Match: 69.09% Indels: 8
DB: 11 Gaps: 3
US-09-890-549-4 (1-759) x AK028186 (1-3100)
QY 56 AsnThrGluAsnLeuSerGlnHisPheArgGlyThrLeuThrValLeuLysLysLys 75
Db 133 AACCTGAAATCTGCCCCAGCACTTTAGAGAGGACCCCTGCTGTGTTAAAGAAAG 192
QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
Db 193 TGGAGAAACCCGGTGGCTGGGGGAGAAATTCACACAGACTCACTGCCAAACACGACG 252
QY 96 GluIleArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly 115
Db 253 GAGGTGGGCACACACGGGACTACCTCTGCTGAAGTACGACGACAGCCCTGCTCTGGA 312
QY 116 AlalysAlaaspGlnGluGlnIleHisProArgSerArgLeuArgSerProProGlu 135
Db 313 GTCAGAGCTACCCGGAAGACGACACCCAGCCCAACCTAGATTGGAATTCGTCGGAA 372
QY 136 AlaleuValGlnGlyArgTyProHisIleLysAspGlyGluAspLeuLysAspHisSer 155

BASE COUNT

908 a 759 c 824 g 609 t
ORIGIN
Alignment Scores:
Pred. No.: 9.01e-200 Length: 3100
Score: 2713.00 Matches: 526
Percent Similarity: 82.55% Conservative: 56
Best Local Similarity: 74.61% Mismatches: 115
Query Match: 69.09% Indels: 8
DB: 11 Gaps: 3

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3100)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tanaka, T., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Db 373 GCAGTTATCCAGAGCGGTATCTCGCTCAGAGAACAGCCACGATTTAAAGCCAGGCC 432
Qy 156 ThrGluSerLysLysMetGluAsnCysLeuGlyCysArgHisGluValGluLysSer 175
Db 433 ACCGAGAGCCAAATAATGGAATACTGTCTGGAGATTCACAGCATTAAGCAGAGAAGCCA 492
Qy 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195
Db 493 GAGACGAGCCAAACACAGAACTTCAGGCAAAATAGAGAAATACACGTTCCACTGAAT 552
Qy 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
Db 553 AGACTGAAGATGATGTTTGAGAAAGGTGAACACAAACCAAGAGTCTCTGGACCCAA 612
Qy 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGlu 235
Db 613 AGCCGAATCGCGGTGGAAGAGGCTCTCTGAAAACAATGTTCCCTGGATCACTGGGAA 672
Qy 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
Db 673 ATAGGTGCGGTCATTGTTCTCTCTGCTGATTCACCTCGAGAAAACAGAGTAAGAG 732
Qy 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
Db 733 AATCTGGAGTGCACGCGCTCTCAGAAACCTCATAAAGGACCGCATGSCCAAGTACCAG 792
Qy 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 793 GCTGCAAGTGTCCAGCAGAGCAGCCAGCCAGCTACACAAATGAGCTGAAACACAGT... 849
Qy 296 GlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluVal 315
Db 850 ---GAAAGCAAACTCATAAATGGGACAGAGGAGATGTCCCCCGAGTCCCCGAGGCC 906
Qy 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluSerLeuAlaValArg 335
Db 907 TGCAGCGTCCATCAGAGCAAGCAAGGTTTCTACAACTGAGAAATAGTCTAGTGGCCCTC 966
Qy 336 SerThrProAlaGluAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 967 TGTGTCCTCGTGAAGATGACACCTGTAACTCCAGGTGAAGAGCAGGCCAGCAGCCT 1026
Qy 356 ValHisProLysProLysSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1027 ATGCACCTTAAGCCGCTAGTCTCTGACGCCAGGACCTCCAGTCTCCGGAAGTTCTCCT 1086
Qy 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1087 TCCAAACACAGCAGAGAGTTTCAGGCGCGCGCAAGAGAGAGCTGCTGGAGTGTCAAGAG 1146
Qy 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPhe 415
Db 1147 ACGGTGTACCCCATGGAACCGGCTCTGGCCAAACAGCAGAGGTGTTTCATCAGCTGTTTC 1206
Qy 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1207 CGATGCTCTACTGCAACCAACAGCTCAGTCTAGGAACATATGCATCTTACATGGACGA 1266
Qy 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1267 ATCTACTGTAGCCTCACTTCAATCACTTCTCAAGTCCAAAGGCAACTATCAGCAGGCG 1326
Qy 456 PheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeu 475
Db 1327 TTTGGCCAAAGCAGCATAAGATCTGTGGCAGCAGAGCGACATAGAGGACACTTTG 1386
Qy 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
Db 1387 GGGAGACAGCCCGCCGCTCAATGAGGGGAGAGCCCGCCATAGCCCGGTGTAGAGAT 1446
Qy 496 AlaProIleAlaLysValGlyValIleAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1447 GCCCCCATCGCAAGGTGCGGCTGTGCGCGCAAGTATGGAAGCCAAAGGCGCTCTCTCAG 1506

Qy 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProPro 535
Db 1507 AGGGAGAGGGAGGATTAAGCTCGGAGACCAAGAAGCTGAGGATCGCTGCTCCGCCA 1566
Qy 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLys 555
Db 1567 GCCGAGCTGGCGGTTCCGGAAGTGCCCTGGAGAGAGGATCAAAAGTATCGAAGCCCAAG 1626
Qy 556 TrpProProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1627 TGGCTCCGAGGATGAGCTGCAAGACGAGAGGCCCGGAGGATGTAGATCTGGACCTG 1686
Qy 576 LysLysLeuArgArgSerSerSerLysLysGluArgSerArgProPheThrValAlaAla 595
Db 1687 AAGAAGCTGCGCGGTCTTCGTCCTGAAGAGAGAGGAGCCCGCTTTACGTAGCAGCT 1746
Qy 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgLysGly 615
Db 1747 TCCCTTGAACCTCTTCATCAAGAGGCCCAAGGCCCTCTTCGCCATCTCTCAGAAAGGT 1806
Qy 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634
Db 1807 TGGAGCAGTCCGAGCAGAGTGAAGGTTTGGAGAGGGATAGCGACGATGGAAGGAAA 1866
Qy 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1867 CAACGGAAAATGCCAGACCTCTCTGGGAGAAAGAGAAATGTGGAAAATCACGCTGGCAG 1926
Qy 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlyHisSerLeuGluMet 674
Db 1927 -----GCCGAGGAAGTCCGAGGAGCAAGGACCGTAGTCTCTTTTGACCTG 1971
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 1972 GAGAGTGAAGATTTATGGAATAATGGAGCAACATAGCTGAAGATGACAAACCATGTCCAC 2031
Qy 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2032 GCACAGCAGTCCCGTTAGAACCCGAGGCTCCGGTTGGTCTTGCTTTGTAGACACACG 2091
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2092 GCTGCTAAAGATTTCACTACCCACATCAAAATCCCAAGATGTCGGGTTCTGGGAGGGA 2151
Qy 735 GluValValLysGluLeuSerValGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2152 GAAGTGTCCGAGAGTGTCTGTGGAGAAACAGATTAAGAGAAACCGGTACTATGATGAG 2211
Qy 755 AspGluAspGluGlu 759
Db 2212 GACGAGGATGAAGAA 2226

RESULT 4
AK085065 3160 bp mRNA linear HTC 05-DEC-2002
LOCUS
DEFINITION
Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched
library, clone:D430032I03 product:epithelial protein lost in
neoplasm, full insert sequence.
ACCESSION
VERSION AK085065.1 GI:26351432
KEYWORDS
SOURCE HTC; CAP trapper.
ORGANISM
Mus musculus (house mouse)
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2

AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
	Genome Res. 10 (10), 1617-1630 (2000)	
	20499374	11042159
JOURNAL MEDLINE PUBMED	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
	Genome Res. 10 (11), 1757-1771 (2000)	
	20530913	11076861
JOURNAL MEDLINE PUBMED	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Asburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kueth, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schrim, L.M., Staubli, F., Suzuki, R., Tonita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Anono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustigich, S., Hill, D., Hofmann, D.A., Hume, D.A., Kaniya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.	
	Functional annotation of a full-length mouse cDNA collection	
	Nature 409 (6821), 685-690 (2001)	
	21085660	11217851
JOURNAL MEDLINE PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
	Nature 420, 563-573 (2002)	
	6 (bases 1 to 3160)	
JOURNAL MEDLINE PUBMED	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
	Direct Submission	
	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken	

FEATURES	source	
	1. .3160	
	/organism="Mus musculus"	
	/mol_type="rRNA"	
CDS	/strain="C57BL/6J"	
	/db_xref="FANTOM DB:D430032103"	
	/db_xref="taxon:10090"	
	/clone="D430032103"	
BASE COUNT	/tissue_type="lung"	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
	/dev_stage="13 days embryo"	
	501. .2282	
ORIGIN	/note="unnamed protein product; epithelial protein lost in neoplasm (MGI:1920992, GB NM_023063, evidence: BLASTN, 99%, match=3086)"	
	putative	
	/codon_start=1	
	/protein_id="BAC39353.1"	
ALIGNMENT SCORES	/db_xref="GI:26351433"	
	/translation="MENCIGSRHEAEKPTSENTSTSGKIEKYNVPLNRLKQMPFKG EHNQTKSWTQSRNAGRRLLSNCSDUWEIGAGHJSSAFNSEKSNLELPRL SKTSIKDRNMQYAAVSKQSPASYTNELKTSKTHKEQKENVPGPEACSVHQEG KVFAPAKSCVCEQKTVYMERLLANQOVFIHSCRCYCNNKLSLGYASLHGRIYC KPHFNOLFKSGVNDSEFGKHQDLWASKDNEETLGRPAQPNAGSPSPGVEDA PIKQGVLAASMEAKASQREDEKPAETKKLRIAMPPELGGSGSALEEGIKVSKP KWPEDDDQTEAPEVDLDDKLRSSSLKERSRPFTVAASFTSSIKSPKASPSFL KRWSSDSQSBEGFGGTATMERKQTNARPSGKENVKSGWQGEVPRSDRSFEL ESENFNGANIADDNHVAQSPLEPPEAPGWSGFVDTTAAKFTTQNKSQDVGFW EGEVVRSLVVEEQIKRNRYYDEDEE"	
	914 a	777 c 843 g 526 t
QY	2.69e-199	Length: 3160
	2707.00	Matches: 525
	82.41%	Conservative: 56
	74.47%	Mismatches: 116
DB	68.93%	Indels: 8
	11	Gaps: 3
QY	US-09-890-549-4 (1-759) x AK085065 (1-3160)	
	56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLys 75	
	186 AACCTGAAGTCTGCCCGACACTTTAGAGAGGGACCTGTCTGTGTTAAAGAGAAG 245	
	76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95	
QY	246 TGGGAGAACCCGGTGGTGGGGCAGAAATCCACACAGACTCCTCCCAACAGCAGCAGT 305	
	96 GlutIleArgHisArgAlaAspHisProAlaGluValThrSerHisAlaIleSerGly 115	
	306 GAGGGTGGCACACAGCGGACTACCTTCCTGCTGAAGTACGGACAGACCTGCTCTCGGA 365	
	116 AlalysAlaAspGlnGluGlnIleHisProArgSerArgLeuArgSerProProGlu 135	
QY	366 GTCAGAGTGACCGGAAGAGCACACCCGCCCAACTAGATTGTGATCTCGTCCGGAA 425	
	136 AlalEuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSer 155	
	426 GCAGTTATCCAGAGCCGGTATCTCTCGTCAGAACACAGCCAGCATTTTAAAGCCAGGCC 485	
	156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175	
QY	486 ACCGAGAGCCCAAAATGGAAACTGTCTGGGAGATTCAGGATGAAGCAGAGACCA 545	

QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysThrAsnValProLeuAsn 195
 DB 546 GAGACAGGCGAAACACAGAACTTCAGGCAAAATAGAGAAATACAACTTCACCTGAAT 605
 QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
 DB 606 AGACTCAAGATGATTTTGAGAAAGGTGAACACCAACCAAGAGTCTCTGGACCCAA 665
 QY 216 SerArgSerAlaSerGlyArgIleSerGluAsnSerTyrSerLeuAspAspLeuGlu 235
 DB 666 AGCGGAAATCGGGTGAAGAGGCTCTCTGAAAACAACTGTCTCCGTGATCAGTGGGAA 725
 QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGlnSerArgArg 255
 DB 726 ATAGGTGCGGGTCATTTGCTCTGCACTCAACTCGGAGAAACAGAGTAAAGAG 785
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
 DB 786 AATCTGAGCTCCACGCTCTCAGAACTCTCAATAAGAGCCGATGCGCAAGTACCAG 845
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
 DB 846 GTGTCAGTGTCCAAAGCAGAGCAGCCAGCCAGCTACACAAATGAGCTGAAACCACT 902
 QY 296 GlyGluLysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluVal 315
 DB 903 ---GAAAGCAAACTCATAAATGGGAACAGAGGAATGTGCCCCAGTCCCGAGGCC 959
 QY 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
 DB 960 TGCAGGCTCCATCAGGAAGAGCAAGGTTTCTCAACTGAGATAGTCTAGTGGCCCTC 1019
 QY 336 SerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
 DB 1020 TCTGTCCCTGCTGAAGTACACCTGTAACTCCAGGTGAAGAGCGAGCCAGGAGCCT 1079
 QY 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
 DB 1080 ATGCACCTTAAGCCGTAGTCTCAGCCAGCAGCCAGCTTCCGGAAGTTCTCT 1139
 QY 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
 DB 1140 TCCAAACACGGAAGATTTTCAGCGCGCGGAAAGGAGAGCTGCGTGGAGTGTGAGAAG 1199
 QY 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPhe 415
 DB 1200 ACGGTGTACCCCATGGAACGGCTCTGCGCAACAGCAGAGGTGTTTCACTCAGCTGTTTC 1259
 QY 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
 DB 1260 CGATGCTCTACTGCAACAAACAGCTCAGTCTAGGAACATATGCATCTTACATGGACGA 1319
 QY 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
 DB 1320 ATCTACTGTAGCCCTCACTTCAATCACTTTCAGTCCAAAGGCAACTATCAGCAGAGC 1379
 QY 456 PheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluLysLeu 475
 DB 1380 TTTGGGCACAGCACATAGGATTTGGGCAAGCAAGAGCGCAATAGGAGACATTTG 1439
 QY 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
 DB 1440 GGGAGACAGCCAGCCCTAATCAGCGGAGAGCCCAATAGCCGGGTGTAGAGAT 1499
 QY 496 AlaProIleAlaLysValGlyValLeuAlaLaserMetGluAlaLysAlaSerSerGln 515
 DB 1500 GCCCCATCGCAGGTGCGGTGCTGCGGCAAGTATGGAAGCAAGGCCCTCTTCTCAG 1559
 QY 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProPro 535
 DB 1560 AGGGAGAGGGAGGATAGCCCTCGGAGACCAAGAGCTGAGGATCGCGTCCCTCCGCCA 1619
 QY 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLys 555

DB 1620 GCCGAGCTGGCGGTTCCGGAAGTCCCTCGAGGAAGGGATCAAAGTATCGAACCCCAAG 1679
 QY 556 TrpProProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
 DB 1680 TGGCTCTCGGAGGATGAGCTCTGCAACAGCAGGAGCCCGGAGGATGATGATCTCGACCTG 1739
 QY 576 LysLysLeuArgArgSerSerSerLeuLysGluLysArgSerArgProPheThrValAlaAla 595
 DB 1740 AAGAAGCTGGCGGTCTTCGTCTCCTGAAGAGAGGAGCGCCCGTTCACGTTAGCAGCT 1799
 QY 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProIleLysGly 615
 DB 1800 TCCTTTTGAACCTCTCCATCAGAGGCCCAAGGCCCTCTTCGCCATCTCTCAGGAAAGGT 1859
 QY 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634
 DB 1860 TGGAGCAGCTCCGAGCAGAGTGGAGGTTTGGAGGAGGGATAGCGACGATGGAAGGAAA 1919
 QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
 DB 1920 CAACGGGAAATGTCACACCTCTCGGAGAAAGAGAAATGTGGGNAATCAGCTGGCAG 1979
 QY 655 AsnLysGluSerLysGlyThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
 DB 1980 -----GCGGAGGAGTCCCGAGGAGCAAGGAGCGCTCTTTTGGAGCTG 2024
 QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
 DB 2025 GAGAGTGAGAAATTTATGGAATGAGCAACATAGCTGAAGATGACAAACCATGTCCAC 2084
 QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
 DB 2085 GCACAGCAGTCCCGTTAGAACCCGAGGCTCCGCGTTGGTCTGCTTTGTAGACACACG 2144
 QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
 DB 2145 GCTGCTAAAGATTCATCTACCCAGATCAAAATCCCAAGATGTCGGGTTCTGGAGGGA 2204
 QY 735 GluValValLysGluLeuSerValGluGluGlnLysArgAsnArgTyrTyrAspGlu 754
 DB 2205 GAAGTGTCTCGAGAGCTCTCTGTGGAGAACACAGATAAAGAGAACCCGTACTATGATGAG 2264
 QY 755 AspGluAspGluGlu 759
 DB 2265 GACGAGATGAAGAA 2279
 RESULT 5
 BM909158
 LOCUS
 DEFINITION BM909158 1075 bp mRNA linear EST 12-MAR-2002
 5', mRNA sequence.
 ACCESSION BM909158
 VERSION BM909158.1 GI:19359537
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1075)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1904 row: j column: 10

High quality sequence stop: 696.

FEATURES

source

1..1075
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5431785"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_98"
 /notes="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 Note: this is a NIH MGC Library."

BASE COUNT 354 a 256 c 258 g 207 t

ORIGIN

Alignment Scores:

Pred. No.: 8,7e-109 Length: 1075
 Score: 1532.50 Matches: 312
 Percent Similarity: 95.78% Conservative: 6
 Best Local Similarity: 93.98% Mismatches: 11
 Query Match: 39.02% Indels: 4
 DB: 12 Gaps: 1

US-09-890-549-4 (1-759) x BM909158 (1-1075)

QY 80 GlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluLeuArgHis 99
 DB 1 GGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGCAGCTGAGATTAGGCAC 60
 QY 100 ArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAsp 119
 DB 61 AGAGCAGACCATCTCTCTGCTGAAGTGAACAGCCAGCTGCTTCTGGAGCCAAAGCTGAC 120
 QY 120 GlnGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGln 139
 DB 121 CAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAGCCCTCTTCTAG 180
 QY 140 GluArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLys 159
 DB 181 GGTCTGATATCCCCACATCAAGACCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAA 240
 QY 160 LysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLeuSerGlu 179
 DB 241 AAAATGGAAAATTTCTAGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAA 300
 QY 180 AsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMet 199
 DB 301 AACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAACAGGCTTAAGATG 360
 QY 200 MetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAla 219
 DB 361 ATGTTTGAAAAGGTGAACCAACTCAACTAAGATTCTCCGGGCCCAAGCCGAAGTGCA 420
 QY 220 SerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluLeuGlyProGly 239
 DB 421 ACTGAAGGAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCAGGT 480
 QY 240 GlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeu 259
 DB 481 CAGTTGTCTATCTTCTACATTGACTCGGAGAAAATAGAGTAGACGAAATCTCGAACTT 540
 QY 260 ProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSer 279
 DB 541 CCACGCCCTCTCAGAAACCTCTATAAAGATCGAATGGCCAAAGTACCAGGACGTGTGTCC 600
 QY 280 LysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluLeuLys 299

DB 601 AAAAAGACGCTCAACCAACTATACAAATGAGCTGAAAGCGAGTGGTGGCGAAATCAAA 660
 QY 300 IleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHis 319
 DB 661 ATTCAATAATGGAGCAAAAGAGATGTGCCCCAGGTCTCTGAGGTCTGATACCCCAT 720
 QY 320 GlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAla 339
 DB 721 CAGGAAGGGGAAAGATTCTGCAATAGCAATAGCTGGCAGTCCGTTCCACCCCTGCC 780
 QY 340 GluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359
 DB 781 GAAGATGACTCCCGTACTCCAGGTTAAGAGTCAAGGTTCAACAGCCTGTCCATCCCAGG 840
 QY 360 ProLeuSerProAspSerArg-AlaSerSerLeuSerGluSerSerProProLysAlaMe 379
 DB 841 CCACTAAGTCAGATTCAGAAAGCTTCAGTCTTCTGAAAGTCTCTCTCCCAAGA-AT 899
 QY 379 tLysLysPheGlnAlaProAlaArgGluThrCys---ValGluCysGlnLysThrValTy 398
 DB 900 GGAAGATTTCGGGACCTGCAGAAAGACCTGGGGGGAATGGCCAAAAAACACTCTA 959
 QY 398 rProMetGlu-ArgLeuLeuAlaAsnGlnGln 408
 DB 960 TCCCATGGAGCGTCTCTTTGGCAACAAGAA 991

RESULT 6

BM914155

LOCUS

DEFINITION BM914155 1142 bp mRNA linear EST 12-MAR-2002
 AGENCOURT_6608184 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5479040
 5', mRNA sequence.

ACCESSION BM914155

VERSION BM914155.1 GI:19364534

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1142)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1999 row: k column: 09
 High quality sequence stop: 678.

FEATURES

source

1..1142
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5479040"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_98"
 /notes="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 Note: this is a NIH MGC Library."

BASE COUNT 376 a 282 c 276 g 208 t

ORIGIN

Alignment Scores:
Pred. No.: 1.49e-98 Length: 1142
Score: 1401.00 Matches: 307
Percent Similarity: 89.24% Conservative: 8
Best Local Similarity: 86.97% Mismatches: 14
Query Match: 35.68% Indels: 24
DB: 12 Gaps: 4

US-09-890-549-4 (1-759) x BM914155 (1-1142)

QY 71 ValLeuLysLysLysTrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeu 90
DB 1 GTGTTAAAGAAAGTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTA 60

QY 91 ArgAsnSerSerThrGluLeuArgHisArgAlaAspHisProAlaGluValThrSer 110
DB 61 CGGAACAGCAGCACTCAGATTAGGCACAGCAGACCATCTCTCTGCTGAAGTGACAAGC 120

QY 111 HisAlaAlaSerGlyAlaLysAlaAspGlnGluGlnIleHisProArgSerArgLeu 130
DB 121 CAGCGTCTCTTGGAGCCAAAGTGACCAAGAAAGAACCAATCCACCAGCATCTAGATC 180

QY 131 ArgSerProProGluAlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAsp 150
DB 181 AGGTCACTCTCTGAAGCCCTCGTTCCAGGTCGATATCCCCACATCAAGCAGGTGAGCAT 240

QY 151 LeuLysAspHisSerThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHis 170
DB 241 CTTAAGACCACTCAACAGAAAGTAAAAAATGGAATACTGTAGGAGAAATCCAGGCAT 300

QY 171 GluValGluLysSerGluLeuSerGluAsnThrAspAlaSerGlyLysIleGluLysTyr 190
DB 301 GAAGTAGAAAAATCAGAAATCAGTCAAAACACAGATGCTTGGGCAAAATAGAGAAATAT 360

QY 191 AsnValProLeuAsnArgLeuLysMetPheGluLysGlyGluProThrGlnThrLys 210
DB 361 AATGTTCCGCTCAACAGGCTTAAGATGATGTTTGAGAAAGTGAAACCACTCAACTAAG 420

QY 211 IleLeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSer 230
DB 421 ATTCTCCGGGCCCAAGCGAAGTCAAGTGAAGGAGAGATCTCTGAAACCTCTATAAGGATCGA 480

QY 231 LeuAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLys 250
DB 481 CTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTCTACATTTGACTCGGAGAA 540

QY 251 AsnGluSerArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArg 270
DB 541 AATCAGAGTAGACCAAAATCTGAACTTCCAGCCTCTCAGAAACCTCTATAAGGATCGA 600

QY 271 MetAlaLysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGlu 290
DB 601 ATGCCCAAGTACCAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAG 660

QY 291 LeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAsnValPro 310
DB 661 CTGAAACCAAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCC 720

QY 311 ProGly-ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAs 330
DB 721 CCAGGTCCCTGAGTCTGCTCATCACCATCAGGAAGGGGAAAGAAATTTCCGCCAATCAGAA 780

QY 330 nSer-LeuAlaValArg-SerThrProAlaGlu-AspAspSerArgAspSer-GlnVal 349
DB 781 TAGCCCTGGCAGTCCGTTTCCACCCCTCGCGAAATATGACTCCCGGAGTCTCCCGAGTTA 840

QY 349 YsSerGlu-ValGlnGln-ProValHisProLys-ProLeuSerProAsp-SerArgAl 367
DB 841 AAGGAGAGGGTTCAACAGGCTCGGCATCCCAAGGCCCTTAAGTCCAGATTTCCAAAGCC 900

QY 367 aSerSerLeuSer---GluSerSerProProLysAlaMetLysLys---PheGlnAlaPr 385
DB 901 CTCAGGCTTTTGAAGGTTCTCCCTCCCAAGCCAAATGAAAAAGGTTTTCGCGACC 960

QY 385 O---AlahrgGluThrCysValGluCysGlnLysThrValTyrPrometGluArgLeuLe 404
DB 961 CTGGAATAAAACCTGC-----CGGGGAAT 987

QY 404 uAlaasnGlnGlnValPhe 410
DB 988 GGCACAAAAACAAGTCTTT 1006

RESULT 7
LOCUS BE410108
DEFINITION 601302288F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637100 5',
mRNA sequence.
ACCESSION BE410108
VERSION BE410108.1 GI:9346558
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 971)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC335 row: o column: 21
High quality sequence start: 56
High quality sequence stop: 806
Location/Qualifiers
FEATURES
source 1..971
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3637100"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_21"
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected 500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 290 a 236 c 243 g 201 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2.41e-90 Length: 971
Score: 1294.00 Matches: 268
Percent Similarity: 85.65% Conservative: 11
Best Local Similarity: 82.45% Mismatches: 33
Query Match: 12.95% Indels: 14
DB: 10 Gaps: 6

US-09-890-549-4 (1-759) x BE410108 (1-971)

QY 269 AspArgMetAlaLysTyrGlnAlaValSerLysGlnSerSerThrAsnTyrThr 288
DB 28 AATCGCGCAGGCGCCAGTCCAGGAGCTGTGTCAAAACAAAGCAGCTCAACCACTATACA 87

QY 289 AsnGluLeuLysAlaSerGlyGluIleLysIleHisLysMetGluGlnLysGluAsn 308
DB 88 AATGAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAAAT 417

309 ValProProGluValCysIleThrHisGlnGluGlyGluValSerAlaAan 328
148 GTCCCCCAGCTCTGAGGTCTGCATCACCATCAGAGGGGAAAGATTCTCGAAT 207
329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
208 GAGAAATAGCTGCGAGTCGCTCCACCCCTGCCGAGAGATGACTCCGTCGACTCCAGGTT 267
349 LysSerGluValGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
268 AAGAGTGAGGTTCAACAGAGCTGTCATCCCAAGCCCACTAAGTCCAGATTCAGAGCCTCC 327
369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
328 AGTCTTCTGAAGTTCTCTCCCAAGCAATGAAGAGTTTTCAGGCACCTCAAGAGAG 387
389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGln 408
388 ACCTGCGTGAATGTGAGAGACAGCTATCAATGAGCGTCTCTTGCCCAACCCAGCAG 447
409 ValPheHisIleSerCysPheArgCysSerTyCysAsnLysLeuSerLeuGlyThr 428
448 GTGTTTTCATCATCAGCTGCTTCGTTGCTCTATTGCAACCACTCAGTCTAGGAACA 507
429 TyrAlaSerLeuHisGlyArgIleTyCysLysProHisPheAsnGlnLeuPheLysSer 448
508 TATGCAATCTTTCATGGAAGATCTATTGTAAGCTCCTCAATCAACTCTTTAAATCT 567
449 LysGlyAsnTyAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468
568 AAGGGCACTATGATGAAGCTTTGGGCACACACACAGAGATCTATGGCAAGCAAA 627
469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
628 AATGAAACGAAAGAGATTGAGAGAGACAGCCAGCTTGCAATGCAAGGAGACCCCT 687
489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaSer--- 507
688 CACAG-CCAGGGGTGAGATGCCCTTATGTAAGTGGGTGCTGGGTGCAAGTAT 746
508 MetGluAlaLysAlaSer-SerGlnGlnGluLysGluAspLysProAlaGluThrLysLy 527
747 TTGGAAGCCCAAGCTCTGCGACCCCACTGAACTTTGGAAGTTCAGGA---GTGCTGGAGGA 863
527 sleuArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGl 547
807 GCTGAAGATGCGCTTGGCACCCCACTGAACTTTGGAAGTTCAGGA---GTGCTGGAGGA 863
547 GglyIleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluVa 567
864 AGGGTCAATGTTAAC---CCCAATGCTCTGAAAGAGAA-----TAGCAGCCCGAGTT 914
567 lProGluAspValAspLeuLysLysLysLeuArgArgSerSerSerLysLysGluAr 587
915 C-----CGGGATGTCTAAATTTGAGAGAAACAACTTTCTCGGGAGAA-- 957
587 gSerArgProPhe 591
958 ---AAGCCCTTT 966
RESULT 8
BM552304
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
BM552304
VERSION
BM552304.1 GI:18790057
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 989)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LiCM1994 row: d column: 22
High quality sequence stop: 666.
Location/Qualifiers
1..989
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5476965"
/tissue_type="astrocytoma grade IV, cell line"
/lab_hosts="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_98"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 339 a 219 c 233 g 197 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 4.09e-87 Length: 989
Score: 1252.50 Matches: 257
Percent Similarity: 93.57% Conservative: 5
Best Local Similarity: 91.79% Mismatches: 13
Query Match: 31.89% Indels: 5
DB: 12 Gaps: 3
US-09-890-549-4 (1-759) x BM552304 (1-989)
QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
Db 130 AACACCGAAATCTCTCCAGCACTTTAGAAAGGGACCTGACTGTGTAAAGAGAAG 189
QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
Db 190 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCACT 249
QY 96 GluIleArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly 115
Db 250 GAGATTAGCACACAGACAGACCATCTCTGCTGAAGTGACAGCCACCTGCTCTCGA 309
QY 116 AlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProGlu 135
Db 310 GCCAAGCTGACCAAGAGAACAATCCACCACCAATCTAGACTCAGGTCTCTCTGAA 369
QY 136 AlaLeuValGlnGlyArgTyProHisIleLysAspGlyGluAspLeuLysAspHis 155
Db 370 GGCCTCGTTTCCAGGTGCGATATCCACATCAAGGACGGTGGAGATCTTAAAGACCACTCA 429
QY 156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLys 175
Db 430 ACAGAAAGTAAATAATGAAAAATTTCTAGAGAAATCCAGGCATCCAGGCAATGAA 489
QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyAsnValProLeuAsn 195
Db 490 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 549

QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
Db 550 AGGCTTAAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCAA 609
QY 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerIleLeuAspLeuGlu 235
Db 610 AGCCGAGTGCAGTGAAGGAGATCTCTGAAACAGCTATTCTCTAGATGACCTGAA 669
QY 236 IleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
Db 670 ATAGGCCAGGTGAGTTGTCTCTCTACATTTCGAGGAGAAATAGAGATGACGA 729
QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
Db 730 AATCTGGAACCTCCAGCCCTCTCAGAACCTCTATTAAGGATCGATGCCCAGTACCA 789
QY 276 AlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeu-LysAlaSerG 295
Db 790 GCAGCTGTGTCACAAACAAAGCAGCTCAACCCACTATACAAATGAGTGAAAGCCGGGG 849
QY 295 YGlyGluIleLysIleHisLysMet-GluGlnLysGlu--AsnValProProGlyProG 314
Db 850 GGCGCAAAATCAAAATCTCTAAATGGAGCAAAAGGAAAGGTGGCCCGGAGGTCCGG 909
QY 314 lu---ValCysIleThrHis---GlnGluGlyGluLysIleSerAlaAsnGluAsn 330
Db 910 AAGGTCTGCTTACCCCATAGGAAAGGGGAAATAATATTCTGGGCAAAAT 965
RESULT 9
BG476496 768 bp mRNA linear EST 21-MAR-2001
LOCUS 602522032F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4640502 5',
DEFINITION mRNA sequence.
ACCESSION BG476496
VERSION BG476496.1 GI:13408775
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 768)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTP/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM1405 row: p column: 07
High quality sequence stop: 745.
FEATURES
Location/Qualifiers
1..768
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4640502"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/notes="Organ: Skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 267 a 150 C 207 G 141 T 3 others

ORIGIN

Alignment Scores:

Pred. No.: 1..62e-86 Length: 768
Score: 1243.00 Matches: 251
Percent Similarity: 97.67% Conservativeness: 0
Best Local Similarity: 97.67% Mismatches: 3
Query Match: 31.65% Indels: 3
DB: 10 Gaps: 0

US-09-890-549-4 (1-759) x BG476496 (1-768)

QY 504 LeuAlaLysSerMetGluAlaLysSerGlnGlnLysGluAspLysProAla 523
Db 2 CTGGTGTCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGAGAGAGCAAGACCCAGCT 61
QY 524 GlnThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySerSerGlySer 543
Db 62 GAAACCAAGAGCTGAGGATCGCTGCGCCACCCCTGAACTTGAAGTTCAGGAAGT 121
QY 544 AlaLeuGluGluGlyIleLysMetSerLysProLysTyrProProGluAspGluLysSer 563
Db 122 CCCTTGGAGGAGGATCAAAATGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 181
QY 564 LysProGluValProGluAspValAspLeuAspLeuLysLysLysLysLysLysLysLys 583
Db 182 AAGCCCGAAGTCTCTGAGGATGTCATGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 241
QY 584 LeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerValLys 603
Db 242 CTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 301
QY 604 SerProLysThrValSerProProLysArgLysGlyTyrSerMetSerGluGlnSerGlu 623
Db 302 AGCCCAAAAACTGTGTCCCTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361
QY 624 GluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLys 643
Db 362 GAGTCTGTGGGTGGAAGAGTTCATAGTNTGGAGATGGAGATGAGAAATCTTGAAGAAATGT 421
QY 644 LysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLysGlyGluThrGly 663
Db 422 AAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGG 481
QY 664 LysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGly 683
Db 482 AAGAGAAATGAAGGAGGATCATAGTNTGGAGATGGAGATGAGAAATCTTGAAGAAATGT 541
QY 684 AlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLys 703
Db 542 GCAGACTCCGATGAAGATGATAACAGCTTCTCAACAAACAATCTCCCAAGAACCCAG 601
QY 704 SerLeuAsnTyrSerSerPheValAsp-AsnThrPheAlaGluGluPheThrThrGlnAs 723
Db 602 TCTCTGAATNGTTCAGTTTNGTAGCAAAACACCTTGTGTAAGAAATTCACACTACAGAA 661
QY 723 nGlnLysSerGlnAspValGluLeu-TyrGluGlyGluValValLysGluLeuSerValG 743
Db 662 TCAGAAATCCCGAGGATGTGGAACTCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 721
QY 743 luGluGlnIleLysArgAsnArgTyrTyrAspGlu-AspGluAsp 757
Db 722 AAGAACAGATAAGAGAAATCGGTATTATGATGAGGATGAGGAT 766

RESULT 10

BUS50203

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

BUS50203 918 bp mRNA linear EST 16-OCT-2002
AGENCOURT 10438215 NIH MGC 109 Homo sapiens cDNA clone
IMAGE:6598982 5', mRNA sequence.
BUS50203
BUS50203.1 GI:24035166
EST.
Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 918)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2832 row: k column: 14
 High quality sequence step: 655.
 Location/Qualifiers
 1..918
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6598982"
 /tissue_type="teratocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_109"
 /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH MGC Library."
 BASE COUNT 313 a 207 c 211 g 187 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8.64e-85 Length: 918
 Score: 1222.00 Matches: 249
 Percent Similarity: 98.04% Conservative: 1
 Best Local Similarity: 97.65% Mismatches: 1
 Query Match: 31.12% Indels: 4
 DB: 13 Gaps: 0

US-09-890-549-4 (1-759) x BU850203 (1-918)

QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
 Db 156 AACACCGAATCTCTCCAGCAGCTTTAGAAAGGGGACCTGCTGTGTTAAAGAGAAG 215
 QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
 Db 216 TGGGAGAACCCAGGGCTGGAGCAGAGTCTCACAGAGTCTCTACGGAACAGCAGCACT 275
 QY 96 GluLeuArgHisArgAlaSerHisProAlaGluValThrSerHisAlaAlaSerGly 115
 Db 276 GAGATTAGCAGACAGACAGACCATCTCTCTGCTGAAGTGACAGCCAGCCTCTCTGGA 335
 QY 116 AlaLysAlaAspGlnGluGlnIleHisProArgSerArgLeuArgSerProGlu 135
 Db 336 GCCTGCTTTCAGGTCGATATCCCATCATCAGGAGCGGTGAGGATCTTTAAAGACCACTCA 395
 QY 136 AlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSer 155
 Db 396 GCCTCGTTTCAGGTCGATATCCCATCATCAGGAGCGGTGAGGATCTTTAAAGACCACTCA 455
 QY 156 ThrGluSerLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
 Db 456 ACAGAAAGTAAAAAATGMAAATTTGCTAGGAGATCCAGGATCCAGGATGAGTAAAGAAATCA 515
 QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195

Db 516 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAAC 575
 QY 196 ArgLeuLysMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
 Db 576 AGCTTAAGATGATGTTTGAGAAAGTGAAACCACTCAAACTAAGATTCTCCGGGCCAA 635
 QY 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGlu 235
 Db 636 AGCCGAAGTGCAAGTGAAGGAGATCTCTGAAAACAGCTATCTCTAGATGACCTGGAA 695
 QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGlyLysAsnGluSerArgArg 255
 Db 696 ATAGGCCAGGTCAGTTGTCTTCTACATTTGACTCGAGAGAAAATCAGAGTAGACGA 755
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
 Db 756 AATCTGGAACCTTCCAGGCTCTCAGAACTCTATAAAAGATCGAATGCGCAAGTACCAG 815
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerG 295
 Db 816 GCAGCTGTGTCACCAACAGAGCAGCTCCACCAACTATACAAATGAGTGGAAAGCAGTG 875
 QY 295 lyGly-GluIleLysIleHis-LysMetGluGlnLys 306
 Db 876 GGGGGGAAATCAAATTCATAAAATGGAGCCAAAG 912

RESULT 11
 BU861039
 LOCUS AGENCOURT 10436477 NIH_MGC 126 Homo sapiens cDNA clone
 DEFINITION IMAGE:6652372 5', mRNA sequence.
 ACCESSION BU861039
 VERSION BU861039.1 GI:24046031
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 787)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: NCI
 CDNA Library Preparation: Michael Brownstein Laboratory
 CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2900 row: 1 column: 04
 High quality sequence step: 617.
 Location/Qualifiers
 1..787
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6652372"
 /tissue_type="mixed (pool of 40 RNAs)"
 /lab_host="DH10B (Ti-phage-resistant)"
 /clone_lib="NIH_MGC_126"
 /note="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc);
 Site 2: SfiI (ggccctccggcc); Double-stranded cDNA was
 prepared from a pool of 40 cell line polyA+ RNAs (bladder
 - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
 kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
 ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
 gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
 used in cloning as follows:
 5'-AAGCAGTGTATCAACGACAGTGGCCATTACGGCGGG-3' and

FEATURES
 source

5'-ATTTAGGCGGCGCGCATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-1 kb size fraction (other fractions present in NIH_MGC_127 and NIH_MGC_128). Library created in the laboratory of T. Usdin, M.D., Ph.D. (NIH, NIH). Note: this is a NIH_MGC Library."

269 a	152 c	217 g	148 t	1 others
-------	-------	-------	-------	----------

BASE COUNT
GC=3939
CG=3939

Alignment Scores:

Pred. No.:	1.11e-84	Length:	71
Score:	1215.50	Matches:	2
Percent Similarity:	96.30%	Conservative:	2
Best Local Similarity:	96.12%	Mismatches:	4
Query Match:	31.05%	Indels:	4
DB:	13	Caps:	1

US-09-890-549-4 (1-759) x BU861039 (1-787)

QY	499	AlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLys	518
Db	4	GCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCCAAGCCTCTCTCTCAGCAGAGAAG	63
QY	519	GluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeu	538
Db	64	GAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTTGGCCACCCCTCTGAACTT	123
QY	539	GlySerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProPro	558
Db	124	GGAAAGTTTCAGAAAGTGCTTGGAGAAAGGATCAAAATGTCAAAGCCCAATGGGCTCTCT	183
QY	559	GluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeu	578
Db	184	GAAGACGAATTCACGACGCCGGAAGTTCCTGAGGATGCGATCTAGATCTGAAGAAGCTA	243
QY	579	ArgArgSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGln	598
Db	244	AGACGATCTTCTTCACTGAAGGAAGAGCGCGCCCATTCACGTGTAGCAGCTTCATTCAA	303
QY	599	SerThrSerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMet	618
Db	304	AGCACCTCTGTCAAGAGSCCAAAACCTGTCTCCCACTATCAGAAAGGCTGGAGCATG	363
QY	619	SerGluGlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsn	638
Db	364	TCAGAGCAGAGTGAAGAGTCTGTGGTGCAGAGTTGCAGAAAGGAACAAGTGGAAAT	423
QY	639	AlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSer	658
Db	424	GCCAGGCTTCTTAAGAAAGATGGATGTGGGAAAAACAACTGCAAAACAAAGAATCT	483
QY	659	LysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsn	678
Db	484	AAAGGAGACAGGGAGAGAGATGAAGNAGTCAATGTTTGGAGATGGAGAATGAGAAT	543
QY	679	LeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSer	698
Db	544	CTTGTAGAAAAATGGTCAGACTCCGATGAGATGATATACAGCTTCTCTCAACACACATCT	603
QY	699	ProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGlu	718
Db	604	CCACAAGAACCCCAAGTCTCTGAATTGGTCGAGTTTCTAGACACACACTTTGCTGAAAGA	663
QY	719	PheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValVal	737
Db	664	ATCACTTACTCAGATCAAAAAATCCCAAGATGTGAACTCTGGNAGGGGAGAAAGTGGTC	723
QY	738	LysGluLeu-SerValGluGluGlnIle-LysArgAsnArgTyrTyr	752
Db	724	AAAAAGAGCTTCTCTGTGGAGAACGGATTAAGAGAGAAATCGGTATT	77

RESULT 12

BU603212	REFERENCE
LOCUS	AUTHORS
DEFINITION	TITLE
ACCESSION	JOURNAL
VERSION	COMMENT
KEYWORDS	
SOURCE	
ORGANISM	

EU603212 847 bp mRNA linear EST 20-SEP-2002
AGENCOURT 10013096 NIH_MGC_142 Homo sapiens CDNA clone
IMAGE:6498219 5', mRNA sequence.

BU603212.1 GI:23254971

KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: NCI

CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2680 row: 1 column: 04

```
High quality sequence start: 118
High quality sequence stop: 500.
```

[illegible]

```

source
1. 847
/biocom/quadrafiles
/organism="Homo sapiens"
/mol_type="mRNA"
/db xref="taxon:9606"
/clone="IMAGE:5498219"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (Tl-phage-resistant)"
/clone_lib="NIH_MGC_142"
/site="Vector: pONR-LIB; Site 1: Sf11
Site 2: Sf11 (ggcgctcgcc); Double-s
prepared from a pool of 40 cell line p
- 2%, blood - 33.4%, brain - 5.6%, bre
- 4%, connective tissue - 1.4%, eye - 1%
kidney - 2.2%, liver - 5.7%, lung - 10
ovary - 4%, pharynx - 2.5%, prostate -
gland - 1.3%, and skin - 2.3%. 5' and
used in cloning as follows:
5'-AACGAGTGATCAACGACAGTGGCCATTACGGC
5'-ATTCTAGAGCGAGCGGCGGCATG-dt(30)NN
enriched library was constructed using
Creator SMART kit and size-selected to
size fraction (other fractions present
Library created in the laboratory of M.
NIH). Note: this is a NIH MGC Library."
285 a 203 c 189 g 167 t 13 othe
BASE COUNT

```

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 4.66E-84
Score: 1212.00
Percent Similarity: 90.97%
Best Local Similarity: 89.24%
Query Match: 30.86%
DB: 13

US-09-890-549-4 (1-759) x BU603212 (1-847)

Qy	105	ProAlaGluValThrSerHisAlaIaSerGlyAlaIysAlaAspGlnGluGlnIle	124
Db	4	CCTGCTGAATGACAGCCACGGTGTCTTGGAGCCAAAGCTGACCAAGAAGAACAAATC	63
Qy	125	HisProArgSerArgLeuAArgSerProProGluAlaLeuValGlnGlyArgTyrProHis	144
Db	64	CACCCCAAGATCTAGACTCAGTCCACTCTCTGAAGCCCTCGTTTCAGGGTCGATATCCCAT	123

QY 696 nGlnSerProGlnGluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAl 716
 Db 545 ACAATCTCCACGAACCAAGTCTCTGAATGGTCGAGTTTGTAGACAACACCTTTGC 604
 QY 716 aGluGluPheThrThrGlnAsnGlnLysSerClnAspValGluLeuTrpGluGlyGluVa 736
 Db 605 TGAAGAAATTCATCTACTCAGAAATCCAGGATGTGAATCTGGGAGGAGAGT 664
 QY 736 lValysGluLeuSerValGluGlnGluGlnLysArgAsnArgTyrTyrAspGluAspGl 756
 Db 665 GGTCAAGAGTCTCTGTGGGAAGACCAAGATACGAGAATCGGTATCATGATGAGGATGA 724
 QY 756 uAspGluGlu 759
 Db 725 GGATGAAGAG 734

RESULT 14
 LOCUS BI089829
 DEFINITION 602855072F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996480 5',
 mRNA sequence.
 ACCESSION BI089829
 VERSION BI089829.1 GI:14508159
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 770)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1021 row: p column: 17
 High quality sequence stop: 707.
 Location/Qualifiers
 1..770
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4996480"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_10"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."

BASE COUNT 250 a 194 c 174 g 152 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.85e-83 Length: 770
 Score: 1199.50 Matches: 245
 Percent Similarity: 96.51% Conservative: 4
 Best Local Similarity: 94.96% Mismatches: 5
 Query Match: 30.54% Incels: 6
 Db: 12 Gaps: 0

US-09-890-549-4 (1-759) x BI089829 (1-770)

QY 233 AspLeuGluLeuGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGlu 252

Db 3 GACCTGGAATATAGCCCAAGGTCAGTTGTCTCTTCTACATTTGACTCGGAGAAAAATGAG 62

QY 253 SerArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAla 272
 Db 63 AGTAGACGAATCTGGAACCTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGCC 122
 QY 273 LysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLys 292
 Db 123 AAGTACAGGACGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAA 182
 QY 293 AlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProProGly 312
 Db 183 GCCAGTGGTGGCAATCAAAATTCATAAATGGACCAAAAGAGAAATGTGCCCCAGGT 242
 QY 313 ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeu 332
 Db 243 CCTGAGGTCTGCATCACCATCAGGAAGGGGAAAGATTTCTGCAATGAGAAATACCTTG 302
 QY 333 AlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluVal 352
 Db 303 GCAGTCGGTTCACCCCTCCGGAAGTGAATCCCGTGAATCCAGGTTAAGAGTGAAGTT 362
 QY 353 GlnGlnProValHisProLysProLeuSerProAspSerArgAspSerLeuSerGlu 372
 Db 363 CAACAGCCTGTCATCCCAAGCCCTAAGTCCAGATACCAAGAGCTCCAGCTCTTTCTGAA 422
 QY 373 SerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGlu 392
 Db 423 AGTTCTCTCCCAAGCAATGAAGTTCAGGCACCTTCGAAGAGAGACCTCGCTGGAA 482
 QY 393 CysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIle 412
 Db 483 TGTCAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCACCAGCAGGTGTAGCACATC 542
 QY 413 SerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeu 432
 Db 543 AGCTCTCTCCGTTGCTCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGTCATCTTA 602
 QY 433 HisGlyArg--IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnT 452
 Db 603 CATGCACAGCAATCTATCGTAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGCAACT 662
 QY 452 TyrAsp-GluGlyPheGlyHis-ArgProHisLysAspLeuTrpAlaSerLysAsnGluAs 471
 Db 663 ATGATGCAAGGCTTTGGCCACCACACACCAAGATCTATGGCAAGCAAAAAATGAAA 722
 QY 471 nGluGluLeuGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
 Db 723 CGAAGAGATACG-GAGAGACGAGCCAGAGA-GCCAATGCAAGGGAG 766

RESULT 15
 LOCUS BM449793
 DEFINITION AGENCOURT_6392972 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528511
 5', mRNA sequence.
 ACCESSION BM449793
 VERSION BM449793.1 GI:18498833
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 931)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/PTP
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

BM449793 931 bp mRNA linear EST 05-FEB-2002
 AGENCOURT_6392972 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528511
 5', mRNA sequence.
 ACCESSION BM449793
 VERSION BM449793.1 GI:18498833
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 931)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/PTP
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

BM449793 931 bp mRNA linear EST 05-FEB-2002
 AGENCOURT_6392972 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528511
 5', mRNA sequence.
 ACCESSION BM449793
 VERSION BM449793.1 GI:18498833
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 931)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/PTP
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

BM449793 931 bp mRNA linear EST 05-FEB-2002
 AGENCOURT_6392972 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528511
 5', mRNA sequence.
 ACCESSION BM449793
 VERSION BM449793.1 GI:18498833
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

http://image.llnl.gov
Plate: LAM12205 row: h column: 16
High quality sequence stop: 575.
Location/Qualifiers
1. .931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:528511"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 296 a 267 c 192 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 1.56e-82 Length: 931
Score: 1193.00 Matches: 246
Percent Similarity: 91.64% Conservative: 6
Best Local Similarity: 89.45% Mismatches: 18
Query Match: 30.38% Indels: 5
DB: 12 Gaps: 2
US-09-890-549-4 (1-759) x BM449793 (1-931)

QY	119	AspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuVal	138
Db	1	GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCCACCTCTGAGCCCTCGTT	60
QY	139	GlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisThrGluSer	158
Db	61	CAGGGTCGATATCCCATCATCAGGCGGTGAGGATCTTAAGACCACTCAACAGAAAGT	120
QY	159	LysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSer	178
Db	121	AAAAAATGGAATAATGCTAGGAGATCCAGGCATGAGTAGAATAAATCAGAAATCAGT	180
QY	179	GluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLys	198
Db	181	GAAACACACATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGACAGGCTTAAG	240
QY	199	MetMetPheGlyLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSer	218
Db	241	ATGATGTTTGAGAAAGTGAACCACTCAAACTAAGATTCTCCGGCCCAAGCCGAAGT	300
QY	219	AlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluLeGlyPro	238
Db	301	GCAAGTGAAGAGAGATCTCTGAAACACAGTATTCTTAGATGACCTGAAATAGGCCCA	360
QY	239	GlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGlu	258
Db	361	GGTCAGTTGTCATCTCTACATTTGACTCGAGAAAAATGAGATAGACGAAATCTGGAA	420
QY	259	LeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaVal	278
Db	421	CTTCCACGCCCTCTCAAAACCTCTATAAGAGATCGAATGGCCCAAGTACCAGCAGCTGTG	480
QY	279	SerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyLeuLe	298
Db	481	TCCAAACCAAGCAGCTCAACCACTATCAAAATGAGCTGAAAGCCAGTGTGGGAAATC	540
QY	299	LysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThr	318
Db	541	AAATTCATAAAATGAGCAAAAGAGAGATGTGCCCCCAGGTCTCTGAGGTCTGCATCACC	600
QY	319	HisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrPro	338
Db	601	CATCAGGAAGCGAAAGATTCTGCATCAGAAATAGCCTGGCAGGCCCTCCCTCCCTCCCT	660

QY	339	AlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnProValHisPro	358
Db	661	GCCGAAGATGATTTCCGTGACTTCTATGTTACCACTGAGGTTCACACAGCTGTCCTCCC	720
QY	359	Lys-ProLeuSerProAspSerArgAla---SerSerLeuSerGluSerSerPro--Pro	376
Db	721	AACGCAACTCAGTCCAGACCCCGAGCCCCCTTTCCCGAAAAAGCTCTCCCTCCCA	780
QY	377	LysAlaMetLysLysPhe---GlnAlaProAlaArgGlu	388
Db	781	AGCGAACGGAATAATTCCTCAGGGCACCCGCAAAAAA	819

Search completed: January 7, 2004, 01:45:14
Job time : 3947 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:10:26 ; Search time 110 Seconds
(without alignments)
3045.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVERQIKRNYDEDEEE 759

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh

-O=/cgn2_1/USPTO_spool/US0980549/runat_06012004_094752_19841/app_query.fasta_1.903

-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCI=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -NATRIX=blosum62 -TRANS=humans40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US0980549 @CIGN_1_1_56@runat_06012004_094752_19841 -NCPU=6 -ICPU=3

-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:

5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3927	100.0	2905	4	US-09-620-312D-45
2	606	15.4	413	4	US-09-016-434-680
3	442.5	11.3	710	4	US-09-016-434-701
4	392	10.0	1638	4	US-09-484-970B-20
5	374.5	9.5	654	4	US-09-016-434-994
6	368	9.4	238	4	US-09-016-434-181
7	226.5	5.8	6755	3	US-08-911-999-4
8	226	5.8	4226	4	US-09-620-312D-480
9	223	5.7	6002	4	US-09-345-882-4
10	212.5	5.4	988	4	US-09-282-146-1
11	205.5	5.2	1015	3	US-08-934-627B-1
12	196	5.0	4323	4	US-09-620-312D-134

13	195.5	5.0	5857	4	US-09-220-132-79	Sequence 79, Appl
14	194	4.9	8532	1	US-08-452-655B-1	Sequence 1, Appl
15	194	4.9	8532	3	US-08-450-582-1	Sequence 1, Appl
16	194	4.9	9606	1	US-07-741-940-1	Sequence 1, Appl
17	194	4.9	9606	1	US-08-289-548A-1	Sequence 1, Appl
18	194	4.9	9606	1	US-08-452-654-1	Sequence 1, Appl
19	194	4.9	9606	2	US-08-370-215A-1	Sequence 1, Appl
20	194	4.9	9606	4	US-07-853-913-1	Sequence 1, Appl
21	193	4.9	11236	1	US-08-449-731-1	Sequence 1, Appl
22	192	4.9	6605	1	US-08-769-309A-4	Sequence 1, Appl
23	192	4.9	6605	3	US-08-994-570-4	Sequence 4, Appl
24	192	4.9	6608	4	US-08-250-132-56	Sequence 4, Appl
25	191	4.9	7453	4	US-09-620-312D-248	Sequence 58, Appl
26	191	4.9	7501	4	US-09-620-312D-249	Sequence 249, App
27	190	4.8	6201	2	US-08-790-912-1	Sequence 1, Appl
28	187.5	4.8	11907	3	US-08-061-376-4	Sequence 4, Appl
29	187.5	4.8	14255	1	US-08-320-559-1	Sequence 1, Appl
30	187.5	4.8	14255	1	US-08-327-392-1	Sequence 1, Appl
31	187.5	4.8	14255	1	US-08-306-691B-55	Sequence 55, Appl
32	187.5	4.8	14255	3	US-08-545-860D-1	Sequence 1, Appl
33	187.5	4.8	14255	5	PCT-US94-04496-1	Sequence 1, Appl
34	185.5	4.7	580073	4	US-08-545-528D-1	Sequence 1, Appl
35	184.5	4.7	3115	4	US-09-620-312D-58	Sequence 58, Appl
36	180	4.6	3972	4	US-09-645-456A-6	Sequence 6, Appl
37	180	4.6	3972	4	US-09-425-324A-6	Sequence 6, Appl
38	180	4.6	3972	4	US-09-645-791-6	Sequence 6, Appl
39	179.5	4.6	4967	3	US-09-177-325-12	Sequence 6, Appl
40	179.5	4.6	4967	3	US-09-411-812A-12	Sequence 12, Appl
41	179.5	4.6	4967	4	US-09-590-113-12	Sequence 12, Appl
42	179.5	4.6	6744	1	US-08-119-125A-2	Sequence 2, Appl
43	179	4.6	15213	4	US-08-961-527-26	Sequence 26, Appl
44	178.5	4.5	4215	4	US-09-620-312D-295	Sequence 295, App
45	178	4.5	3807	4	US-09-645-456A-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-09-620-312D-45
; Sequence 45: Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aigong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 78CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ IDS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 45
; LENGTH: 2905
; TYPE: DNA

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (173)...(2452)
 US-09-620-312D-45

Alignment Scores:

Pred. No.: 0 Length: 2905
 Score: 3927.00 Matches: 759
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-09-890-549-4 (1-759) x US-09-620-312D-45 (1-2905)

QY 1 MetGluSerSerProPheAsnArgGlnThrThrSerLeuSerLeuArgValThrAla 20
 Db 173 ATGGAATCATCTCCATTAAATAGACGGCAATGGACCTCACTATCATTTAGGGTAAACAGCC 232
 QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
 Db 233 AAAGAACTTCTCTGTGCAACAGAAACAAGTCATCGGCTATTGTGGAATATTCTCCAG 292
 QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
 Db 293 TACCAGAAACAGCTGAGAAACAACATGGAGAGAGAAAGTAAACCCGAAATCTC 352
 QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysThrGluAsnProGly 80
 Db 353 TCCAGCACATTTAGAAAGGGGACCTGACTGTGTAAAGAGAGAGTGGGAGAACCCAGGG 412
 QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
 Db 413 CTGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGA 472
 QY 101 AlaAspHisProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
 Db 473 GCAGACCATCTCTCTGCTGAGTGAAGTGAACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCA 532
 QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProGluAlaLeuValGlnGly 140
 Db 533 GAAGAAACAATCCACCCAGATCTAGACTCAGTCACTCTCTGAGCCCTCTCTTCAAGGT 592
 QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
 Db 593 CGATATCCCCACATCAAGGACCGTGAGGATCTTAAAGACCCTCAACAGAAAGTAAAGAA 652
 QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 Db 653 ATGGAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAGAAATCAGAAATCAGTGAAAC 712
 QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 Db 713 ACAGATGCTTCGGGCAAAATACAGAAATATAATGTTCCGCTGAAACAGGCTTTAAGATGATG 772
 QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 Db 773 TTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGTGCAGT 832
 QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
 Db 833 GGAAGGAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAATAGGCCAGGTCCAG 892
 QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuLeuPro 260
 Db 893 TTGTCATCTTCTACATTTGACTCGAGAGAAATAGAGTAGACGAAATCTCGAACTTCCA 952
 QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
 Db 953 CGCCTCTCAGAAACCTCTATAAGGATCGAATGCCAAGTACCGAGGACGCTGTGTCCAAA 1012
 QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300

Db 1013 CAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 1072
 QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 Db 1073 CATAAATGGAGCAAAAGAGAAATGTGCCCCAGGTCTCTGAGGTCTGATCATCCCATCAG 1132
 QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 Db 1133 GAAGGGGAAAGATTTCTGCAATAGAAATAGCTGGCAGTCCGTTCACCCCTCCGCGAA 1192
 QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
 Db 1193 GATGACTCCCGTCACTCCAGGTTAAGAGTGAAGTTCACAGCCTGTCTATCCCAAGCCA 1252
 QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
 Db 1253 CTAAGTCCAGATTCAGAGCCTCCAGTCTTCTGAAAGTTCCTCTCCCAAGCAATGAAG 1312
 QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
 Db 1313 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAAGAGACAGTCTATCCAATG 1372
 QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
 Db 1373 GAGCGTCTCTTGGCCCAACACAGCAGGTGTTTCACATCAGCTGCTTCCTGCTCTTATTC 1432
 QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
 Db 1433 AACCAACAACACTCAGTCTAGGAACATATGCATCTTTACATGGAAGAAATCTATTGTAGGCT 1492
 QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
 Db 1493 CACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGCAGACAGACCA 1552
 QY 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
 Db 1553 CACAAGGATCTATGGGACAGCAAAATGAAACGAAGAGATTTTGGAGAGACCAAGCCAG 1612
 QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
 Db 1613 CTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAAAGTGCCTCTATTGCTAAG 1672
 QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysLysGluAsp 520
 Db 1673 GTGGGTCTCTGGCTGCAAGTATGGAAGCAAGCCCTCTCTCAGCAGGAGAGAGAGAC 1732
 QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProProThrGluLeuGlySer 540
 Db 1733 AAGCCAGCTGAAACCAAGAGAGCTGAGGATCGCTGCGCCACCCCTCACTGAAGT 1792
 QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysThrProProGluAsp 560
 Db 1793 TCAGGAAGTGCCTTGGAGGAGAGGATCAAAATGTCAAAGCCCAATGCGCTCTCTGAAGAC 1852
 QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgAtg 580
 Db 1853 GAAATCAGCAAGCCCAAGTCTCTGAGGATGTCATCTAGATCTGAAGAGCTTAAGACGA 1912
 QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
 Db 1913 TCTTCTTCACTGAAGGAAAGAGAGCCGCCATTCCTGAGCAGCTTCATTTCAAAGCACC 1972
 QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSerGlu 620
 Db 1973 TCTGTCAAGAGCCCAAACTGTGTCTCCCACTATCAGAAAGGCTGGAGCATGTGAGAG 2032
 QY 621 GlnSerGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
 Db 2033 CAGAGTGAAGAGTCTCTGGTGGAGAGTTCAGAAAGGAAACAAGTGGAAATGCCAAG 2092
 QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660

Db 2093 GCTTCTTAAGAAGAAATGGGAATGGGAAACAACTGGCAAAACAAAGAAATCTAAAGGA 2152
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 GAGACAGGAAGAGAGAGTAAAGAGGTCTAGTTTGGAGATGGAGATGAGAAATCTTGA 2212
QY 681 GluAsnGlyAlaAspSerLysGluAspAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCAACAAACAATCTCCANA 2272
QY 701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCCAGTCTCTGAAATGGTCGAGTTTGTAGACAACACACTTTGCTGAGAAATCTACT 2332
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGlyValValLysGluLeu 740
Db 2333 ACTCAGAAATCAGAAATCCAGAGTGGAGACTCTGGAGGAGAGAGTGGTCAAAGAGTCT 2392
QY 741 SerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluGluGlu 759
Db 2393 TCTGTGGAAGACAGATAAGAGAAATCGGTATTATGATGAGATGAGATGAGAG 2449

RESULT 2
US-09-016-434-680
; Sequence 680, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 845-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 680:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2NOT01
; CLONE: 269891
US-09-016-434-680

Alignment Scores: 2.16e-50 Length: 413
Pred. No.: 606.00 Matches: 127
Score:

Percent Similarity: 92.09% Conservative: 1
Best Local Similarity: 91.37% Mismatches: 8
Query Match: 15.43% Indels: 4
DB: 4 Gaps: 0

US-09-890-549-4 (1-759) x US-09-016-434-680 (1-413)

QY 296 GlyGluLeuLysLysLysLysMetGlu-GlnLysGluAsnValProGlyProGluVal 315
Db 3 GGCAGAAATCAAAATTCATAAAATGGAGNCAAAAGAGCAATGTGCCCCAGGTCTGAGGT 62
QY 315 lcySrLysThrHisGln-GluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaVal 335
Db 63 CTGCATCACCATCAGGAAAGGAAAGATTTCTGCAAAATAG-AAATAGCTGGCAGTCC 121
QY 335 rgsrThrProAlaGluAspSerArgAsp-SerGlnValLysSerGluValGlnGln 354
Db 122 GTTCCANCCCTGCGAAGATGACGGGGGGAATCTCCANTTTTGTAGTGGAGTTCACAG 181
QY 355 ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer 374
Db 182 CCGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCTCCAGTCTTCTGAAAGTTCT 241
QY 375 ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln 394
Db 242 CCTCCCAAGCAATGAAGAGTTTTCAGGCACCTGCAAGAGAGACCTGCGTGGATGTGAC 301
QY 395 LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCys 414
Db 302 AAGACAGTCTATCCATGGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCCATCAGTGC 361
QY 415 PheArgCysSerTyrCysAsnLysLeuSerLeuGlyThrTyrAlaSer 431
Db 362 TTCCGTGCTCTTATTCGAAACAAACAACTCAGTCTAGGAACATATGCATCT 412

RESULT 3

US-09-016-434-701
; Sequence 701, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 845-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 701:

SEQUENCE CHARACTERISTICS:
; LENGTH: 710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HEARNOT01
; CLONE: 305403
US-09-016-434-701

Alignment Scores:

Pred. No.: 5.18e-34 Length: 710
Score: 442.50 Matches: 101
Percent Similarity: 53.06% Conservative: 29
Best Local Similarity: 41.22% Mismatches: 88
Query Match: 11.27% Indels: 27
DB: 4 Gaps: 5

US-09-890-549-4 (1-759) x US-09-016-434-701 (1-710)

```
QY 387 ArgGluThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsn 406
Db 6 AGGAAATATGTCATACCTTTGTCAAAGRCAGTTTATCCATGGAGTGCCTAGTGGCAGAC 65
QY 407 GlnGlnValPheHisIleSerCysPheArgCysSerTyrcysAsnAsnLysLeuSerLeu 426
Db 66 AAGCAGAAATTTTCATAGTCTGCTCCGATGCCACCATTTGCAACAKTAAACTAAGTTTG 125
QY 427 Gly-ThrTyAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsnGlnLeuPh 446
Db 126 GGTAAATATGTCATACCTTTGTCAAAGRCAGTTTATCCATGGAGTGCCTAGTGGCAGAC 185
QY 446 eLysSerLysGlyAsnTyrcysLysGlnGluGlyPheGlyHisArgProHisLysAspLeuTrpAl 466
Db 186 CAATCCAAAGGAAATATATGATGAAGTTTGTGACATATAAGCAGCATATAAGATAGTGAA 245
QY 466 aSerLysAsnGluAsnGluGluLysLeu-----LeuGluArgProAlaGlnLeuAl 482
Db 246 CTCGAAATAACCAAGCAGATACGTGACTTTATTCCTAATGAGAACCAAAATATGTGTAA 305
QY 482 aAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLysValGl 502
Db 306 AAATATTGCGAGAAACACACCTTTGTCAGATCGTATGACATTTA-----354
QY 502 yValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp----- 520
Db 355 -----GATCGTGTAACTACAGTGAAGGCAAGGAATGATTGAG 392
QY 521 -LysProAlaGluThrLysLysLeuArgIleAlaLysProProProThrGluLeuGlySe 540
Db 393 AAAATTAGGGGAAGGGGAAATTTAAAGTCAATTTGGCCCTCTTCCAAAGGAGATCCCTAA 452
QY 540 rSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTyrProProGlu-- 559
Db 453 GAAACCTTTACCTTTGAGGAAGAGCTCAAAATGATGATTAACCTAAGTGGCCACCTGAAAT 512
QY 560 -AspGluIleSerLysProGluValProGluAspValAspLeu-AspLeuLysLysLeuA 579
Db 513 GACAACTTCCTATCCCTGAAATTTAAAGTCAATTTGCTAGGAGATGTAGGACTCC 572
QY 579 rArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlns 599
Db 573 AGCAATAAAGNCAAGCAAGGANCACCTTTCCATTTTTCGACCTTATCTNCCAGTC 632
QY 599 erThrSerVal-----LysSerProLysThrValSerProProI 612
Db 633 CACCCCATGTTTGNCGAGGAGNGGTTTGTGTTGNGGGTCCAAAGGAATGGAATAGCCCT 692
QY 612 leArgLysGly 615
Db 693 TAAGGGGAGGG 703
```

RESULT 4

US-09-484-970B-20
; Sequence 20, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 1638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 200145-SCBI
; NAME/KEY: unsure
; LOCATION: 1554, 1581, 1624
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-20

Alignment Scores:

Pred. No.: 1.81e-28 Length: 1638
Score: 392.00 Matches: 124
Percent Similarity: 42.60% Conservative: 43
Best Local Similarity: 31.63% Mismatches: 135
Query Match: 9.98% Indels: 93
DB: 4 Gaps: 13

US-09-890-549-4 (1-759) x US-09-484-970B-20 (1-1638)

```
QY 330 AsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnValLys 349
Db 103 AACCAGCGGGTGGCGCTTCTCCACC-----GAGGCTTCCACTCCACAGGCCATGT 156
QY 350 SerGluValGlnGlnProValHisProLysProLeu---SerProAspSerArgAlaSer 368
Db 157 TCCAGGTGAGGAGCGCCAGCCGCCCTCTCATGAGCCCAAGAGGGGGGGGAGC 216
QY 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPhe-----GlnAlaProAla 386
Db 217 AGCAGGTGAGCGCTCC-----AAGTCCTTACGCTGCGGGCCAGGTG 261
QY 387 ArgGluThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsn 406
Db 262 AAGGAGACCTTGGCGCGCTGCCAGAGACCGTGTACCCCATGGAGCGGTGGTGGCGAC 321
QY 407 GlnGlnValPheHisIleSerCysPheArgCysSerTyrcysAsnAsnLysLeuSerLeu 426
Db 322 AAGCTCATTTTCCACAACTCTTCTGCTGCTGCAAGCACTGTCCACCAAGCTCAGCCTG 381
QY 427 GlyThrTyAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsnGlnLeuPhe 446
Db 382 GGAGCTACCGCGCTGCCAGGGGAGTTCTACTGCAAAACCCCACTTCCAGCAGCTGTTT 441
QY 447 LysSerLysGlyAsnTyrcysLysGlnGluGlyPheGlyHisArgProHisLysAspLeuTrpAla 466
Db 442 AAGAGCAAAAGGCAACTTACGACGAGGGGTTTGGCGCGCAAGCAGCACAAGAGCTCTGGCC 501
QY 467 SerLysAsnGluAsnGluGluLysLeuGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
Db 502 CACAAGGAGGTGAC-----516
QY 487 ThrProHisSerProGlyValGluAspAla-----Profile 498
Db 517 -----CCCGCACCAAGCGCTCTGAGGCTCTGTAACTTCCACCCCTCTG 564
QY 499 AlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLys 518
Db 517 -----CCCGCACCAAGCGCTCTGAGGCTCTGTAACTTCCACCCCTCTG 564
```

NAME: Zeller, Karen J.
 REGISTRATION NUMBER: 37,071
 REFERENCE/DOCKET NUMBER: PA-0002 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 994:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 654 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: STOMTUT01
 CLONE: 889949
 US-09-016-434-994

Alignment Scores:
 Pred. No.: 2,01e-27 Length: 654
 Score: 374.50 Matches: 89
 Percent Similarity: 53.46% Conservative: 27
 Best Local Similarity: 41.01% Mismatches: 46
 Query Match: 9.54% Indels: 56
 Gaps: 8

US-09-016-434-994 (1-654)
 565 -----CGGAAGCCCTGGAGCCGGCAGCGGGAA 591
 519 GluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGlu--- 537
 592 GGTGGAGAGGAGTCGAG-----CTGG-CTTGGTGGGGCCAGG-TGGGAAGGG 640
 538 -----LeuGlySerSerGlySerAlaLeu----- 545
 641 GATGAGGCTTCTCAGCGGTAGGAGCAGGAGGCTCTGCTCCAGATCTCTTCTT 700
 546 -----GluGluGlyLeuLysMetSerLysProLysTrpProProGlu 559
 701 CTCCTTCTCCGAGCCGCTGAGGCTTGGAAACACAGGATTGGGCTGCTCCACACCC 760
 560 -----AspGluLeuSerLysProGluValProGluAsp-----Val 571
 761 TGTCTCTGCTTCTGCTCAGCTCCCTCCACCTCACCAGGAGCCCTCGGAGGCCCC 820
 572 AspLeuAspLeuLysLysLeuArgArgSerSerSerLysLysGluArgSerArgProPhe 591
 821 CAAGCCAGCTCCCTATCTAGTGCTCTTCTCCAGCAAGGAGTCAGCATGCCCCCTC 880
 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
 881 AGGTGCTCCAGCTCCCTCAGTCCACCGGAGA-CTGTGTGCCCCCAGCTCTCCCATCT 939
 612 IleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAla 631
 940 ACCCTACCTTAACTCTTCTGAGCCACGAGGAGGAGGAGGCGGACAGTGC 999
 632 GluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsn----- 647
 1000 CACCTGTTGGGATCATATAATCCCTCCTCAGGCCATGGGGAGGAGATGGGAAAGTGAG 1059
 648 -----ValGlyLysThrTrpGluAsnLysGluSerLys 659
 1060 CCACCTGCTCTGAGGCGCAAGGAGGCGCTGCCCTGCTGGGGCTTGGGACCACTCGA 1119

RESULT 5

US-09-016-434-994
 Sequence 994, Application US/09016434
 Patent No. 6500398
 GENERAL INFORMATION:
 APPLICANT: Janice Au-Young
 APPLICANT: Jeffrey J. Seilhamer
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 NUMBER OF SEQUENCES: 1490
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM: disk
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 CURRENT APPLICATION DATA: US/09/016,434
 APPLICATION NUMBER: HEREWITH
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION NUMBER:
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:

RESULT 6

US-09-016-434-181

RESULT 7

Qy	42	GlnLysAlaAlaGluThrAsnMetGluLysLysArgSerAsnThrGluAsnIeuSer	61
Db	3150	AGGAAAGAAGGGGGAACACAAAGAAACGAAAAACAACCGGCAACAGCCCAACCAAAACG	3209
Qy	62	GlnHisPheArg-----LysGlyThrLeuThrValLeuLys-----	73
Db	3210	GCCAGGAAAGACAGAGCGAAAGGGCAAGCAGGAGGCACCAACAAAGAAACAAGACAA	3269
Qy	74	-----LysLysTrpGluAsnProGlyLeuGlyValaGluSerHisThrAspSer	89
Db	3270	AGCCGCAACCCAAAAGAGGAGNACCCACCAAGGCGCACAGAGCAAGCAACACG	3329
Qy	90	LeuArgAsnSerSerThrGluLeuArgHisArgAlaAspHisProAlaGluValThr	109
Db	3330	AGAAAAGACCAACAAACGAAAGCCACGACGAAAGACGAAAAAAGGAGGCACAG	3383

Dd		5234 AAAAACCGGAACCAAGAAAGCAGAACAAGAAAAAAGGAGCGAAAAGAAAAACGGGC	: : : : : :
Oy		758 luGlU 759 :::	:
Dd		5294 AAAG 5298 ::::	:::
RESULT 8			
US-09-620-312D-480			
; Sequence 480, Application US/09620312D			
; Patent No. 6569662			
; GENERAL INFORMATION:			
; APPLICANT: Tang, Y., Tom			
; APPLICANT: Liu, Chenghua			
; APPLICANT: Asundi, Vinod			
; APPLICANT: Zhang, Jie			
; APPLICANT: Ren, Feiyun			
; APPLICANT: Chen, Rui-hong			
; APPLICANT: Zhao, Qing A.			
; APPLICANT: Wehrman, Tom			
; APPLICANT: Xue, Aidong J.			
; APPLICANT: Yang, Yonghong			
; APPLICANT: Wang, Jian-Rui			
; APPLICANT: Zhou, Ping			
; APPLICANT: Ma, Yunqing			
; APPLICANT: Wang, Dunrui			
; APPLICANT: Wang, Zhiwei			
; APPLICANT: John Tillinghurst			
; APPLICANT: Drmanac, Radcoje T.			
; TITLE OF INVENTION: Polypeptides			
; FILE REFERENCE: 784CIP2B			
; CURRENT APPLICATION NUMBER: US/09/620,312D			
; CURRENT FILING DATE: 2000-07-19			
; PRIOR APPLICATION NUMBER: 09/552,317			
; PRIOR FILING DATE: 2000-04-25			
; PRIOR APPLICATION NUMBER: 09/488,725			
; PRIOR FILING DATE: 2000-01-21			
; NUMBER OF SEQ ID NOS: 1105			
; SOFTWARE: pt_FL_genes Version 1.0			
; SEQ ID NO 480			
; LENGTH: 4226			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)..(2982)			
US-09-620-312D-480			
Alignment Scores:			
Pred. No.:	1.5e+ll	Length:	4226
Score:	226.00	Matches:	181
Percent Similarity:	35.39%	Conservative:	146
Best Local Similarity:	19.59%	Mismatches:	319
Query Match:	5.76%	Indels:	278
DB:	4	Gaps:	43
US-09-890-549-4 (1-759) x US-09-620-312D-480 (1-4226)			
OY	40 LysTyrGlnLysAlaalaGluThrAsnMetGluLysLysArgSerAsnThrGluAsn	59	
Ddb	250 GAATTCAGATGCCATTCCAGAGAAAGTTGTATAACAGCAATGTAAGGAGTGTAATAAT	309	
OY	60 LeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnPro	79	
Ddb	310 GTAAAGAA-----ATAAAGTTAAGGAGGAAATGAA	342	
OY	80 GlyLeuGlyAlaGluSerHisThrAspSerIleuArgAsnSerSerThrGluileArghis	99	
Ddb	343 ACAGAGATCCAAGAAATAAGATGAGGAGGAGGAAT-----ATAATACC	390	
OY	100 ArgAlaAspHISproProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaasp	119	

```
QY 374 SerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCys-----Val 391
|||
Db 1489 TCAAAACCA-----CAATAAAACGTTGTAAGAGAGTATTGCAATACAGAA 1536
|||
QY 392 GluCysGlnLysThrValTyPro-----MetGluArgLeuLeuAlaAsnGlnGlnVal 409
|||
Db 1537 GAGTGTCTAAAAACTGATACCTCGGCAAAAGGAGAGAGGCAAGAAACAAAGAA--- 1593
|||
QY 410 PheHisLeSerCysPheArgCysSerTyCys----- 420
|||
Db 1594 -----TCATTGTCATGGAAACACAGTACCAACAGCTCTTCA 1629
|||
QY 420 ----- 420
Db 1630 GATGAAGATGAAGAAGAAACAAAGACCAAGATGACACCAACTAAGAAATACAATGGTTG 1689
|||
QY 421 --AsnAsnLysLeuSerLeuGluThrTyAlaSerLeuHisGlyArgIleTyCysLys 439
|||
Db 1690 GAGGAAAAAGAAAATCTCTACGGCAACTGGTTCTTATTACGAA----- 1734
|||
QY 440 ProHisPheAsnGlnLeu-----PheLysSerLysGlyAsnTyArgGluGly 455
|||
Db 1735 -----TTTTTCAGAGTGGCAGAAAAGAGATTAACTTTTAAATAACTCTGATGAAGA 1788
|||
QY 456 PheGlyHis-----ArgProHisLysAspLeuTrpAlaSer----- 467
|||
Db 1789 CTTCAAAACAGCAGGCGCAAGATCGAAAAGATGTCGTGGTCAAGTATTACGGGACAGTGG 1848
|||
QY 468 --LysAsnGlnAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGlu 486
|||
Db 1849 CCTAAAAAACGCTGAAAAGAGCTTTTTTCA-----GACTCTGATACCTCAGGCTGCAGCTTCC 1905
|||
QY 487 ThrProHisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaLa 506
|||
Db 1906 CCACCGCAT-----CCTGCCCCAGAGAGGGGTGGCAGAGGAG 1944
|||
QY 507 SerMetGluAlaLysAla-----SerSerGlnGlnGluLys--- 518
|||
Db 1945 TCACCTCAGACTGTGCTGAGAGAGAGAGTGTTCACCCAGTGTAGAACTAGAAAAACCA 2004
|||
QY 519 -----GluAspLysProAlaGluThrLysLysLeuArgIle----- 530
|||
Db 2005 CCTCCAGTCAATGTCGATAGTAACCCATTGAAGAAAAACAGTAGAGTCAATACAGAGA 2064
|||
QY 531 -----AlaTrpPro-----ProProThr----- 536
|||
Db 2065 AAGCGAATTTCCAGTAGTCGCAGTAATTCAGTGCCTAAATACCCCTCTACTACACCT 2124
|||
QY 537 GluLeuGlySerSer-----GlySerAlaLeuGluGluGlyLysMet 551
|||
Db 2125 GAATCGCTTCTATCATGCTACTTAACAGAGGCGCGCAGCAGCTTCTGTAAACAGTA 2184
|||
QY 552 SerLysProLysTrpProGluAspGluLysSer-----LysProGluValProGlu 569
|||
Db 2185 TCAGAACCACTGGCTCCAAACCAAGAGAGGTTTCAAGTATCAAGAGTGAACCTGATAGC 2244
|||
QY 570 AspValAspLeuAsp-----LeuLysLysLeuArgArgSerSerSerLeu 584
|||
Db 2245 ACAATTGAGGTGGATAGTGTGCTGGGAGCTCCAAAGACCTCCAGTCGAAAGGAAT--- 2301
|||
QY 585 LysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSer 604
|||
Db 2302 -----ACTGCCACAGAGGTTTGTATCCAGTGTGAGCTCAAGCAGTAGTAATCAG 2352
|||
QY 605 ProLysThrValSerProPheArgLysGlyTrpSerMetSerGluGlnSerGluGlu 624
|||
Db 2353 CCAGAACCAAGATCCT-----GAAAAGCCTGTACAGTCAAGAAAAGAGTGAAGAT 2406
|||
QY 625 SerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAla----- 641
|||
Db 2407 GCTCAGGAGGAGGAAGTTCATCAAAAAGACGAGAAAGACCCATAAAGCAACAGTGGTA 2466
|||
```

```
QY 642 ---SerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
|||
Db 2467 AACACAAAAGAGGAAAGGCAAAAGCACAATAGTAGTAGTGAAGAACTTTTCAGCTGGT 2526
|||
QY 661 GluThrGlyLysArgSerLys-----GluGlyHisSer 671
|||
Db 2527 GAAAGTATTAATAAGAGTCAAGCAGTCAATCAGTTTCCACCTGGGAATGAAGTCTCATAGT 2586
|||
QY 672 LeuGlu-----MetGluAsnGlnLeuValGluAsnGlyAlaAspSerAspGluAsp 689
|||
Db 2587 ACCAATCTCCCGCAGGACGAGCTCCAGGAAAATGTGGAAGAAATGCTGATAAGGAT 2646
|||
QY 690 AspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer--- 708
|||
Db 2647 CTGAT-----CTCAGGAAACCCAGTAATCGATTACCCAAAAGTTTACAAATGGAGTTT 2700
|||
QY 709 -----SerPheValAspAsn---ThrPheAlaGluGluPheThrThrGlnAsnGlnLys 725
|||
Db 2701 CAGATGTCGAGCTCGAAAATATGACAAAGTCCGAAACGATCACAATTTCTTCAAGAAAA 2760
|||
QY 726 SerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu---SerValGluGlu 744
|||
Db 2761 CTTCAAGAAATCAGAAAACATTTATCTGTCATTAAAAATCTGAAGTAGCTTCCATTGATCGG 2820
|||
QY 745 GlnIleLysArg 748
|||
Db 2821 AGGAGAAAGCGT 2832
|||
```

RESULT 9

```
US-09-345-882-4
; Sequence 4, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bouquelier, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345.882
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 6002
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1319
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1338
; OTHER INFORMATION: 5-130-276 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1944
; OTHER INFORMATION: 5-136-174 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 3329
; OTHER INFORMATION: 5-143-84 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 3346
; OTHER INFORMATION: 5-143-101 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 4582
; OTHER INFORMATION: 5-148-352 : polymorphic base G or T
```



```
FEATURE:
NAME/KEY: allele
LOCATION: 1107..1125
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 1107..1125
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 1315..1338
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 1315..1338
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 1921..1967
OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID41
FEATURE:
NAME/KEY: allele
LOCATION: 1921..1967
OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID62
FEATURE:
NAME/KEY: allele
LOCATION: 3306..3352
OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID46
FEATURE:
NAME/KEY: allele
LOCATION: 3306..3352
OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID67
FEATURE:
NAME/KEY: allele
LOCATION: 1296..1338
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 1296..1338
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 3323..3369
OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID45
FEATURE:
NAME/KEY: allele
LOCATION: 3323..3369
OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID66
FEATURE:
NAME/KEY: allele
LOCATION: 4559..4605
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID48
FEATURE:
NAME/KEY: allele
LOCATION: 4559..4605
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID69
FEATURE:
NAME/KEY: misc_feature
LOCATION: 442..444
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4378..4380
OTHER INFORMATION: stop : TGA
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 4878..4883
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 5116..5121
OTHER INFORMATION: potential
```

```
NAME/KEY: polyA_signal
LOCATION: 5896..5901
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 5961..5986
FEATURE:
NAME/KEY: misc_feature
LOCATION: 209..756
OTHER INFORMATION: homology with EST in ref embl:W84531
FEATURE:
NAME/KEY: misc_feature
LOCATION: 391..815
OTHER INFORMATION: complement homology with EST in ref embl:W37603
FEATURE:
NAME/KEY: misc_feature
LOCATION: 453..898
OTHER INFORMATION: complement homology with EST in ref embl:H39516
FEATURE:
NAME/KEY: misc_feature
LOCATION: 818..1306
OTHER INFORMATION: complement homology with EST in ref embl:W67770
FEATURE:
NAME/KEY: misc_feature
LOCATION: 844..1303
OTHER INFORMATION: complement homology with EST in ref embl:AA262427
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1351..1702
OTHER INFORMATION: complement homology with EST in ref embl:AA485189
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1866..2109
OTHER INFORMATION: homology with EST in ref embl:AA296993
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2181..2281
OTHER INFORMATION: homology with EST in ref embl:T61718
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2253..2482
OTHER INFORMATION: homology with EST in ref embl:AA082927
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2480..2842
OTHER INFORMATION: complement homology with EST in ref embl:H38607
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3334..3733
OTHER INFORMATION: homology with EST in ref embl:AA279595
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3631..3870
OTHER INFORMATION: complement homology with EST in ref embl:AA169631
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3883..4221
OTHER INFORMATION: homology with EST in ref embl:H08612
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4277..4796
OTHER INFORMATION: homology with EST in ref embl:AA399016
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4516..5016
OTHER INFORMATION: homology with EST in ref embl:AA479433
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5580..6002
OTHER INFORMATION: complement homology with EST in ref embl:AA167428
US-09-345-882-4
```

Alignment Scores:

Pred. No.: 5,31e-11 Length: 6002
 Score: 223.00 Matches: 178
 Percent Similarity: 35.51% Conservative: 148
 Best Local Similarity: 19.39% Mismatches: 326
 Query Match: 5.68% Indels: 266
 DB: 4 Gaps: 42

US-09-890-549-4 (1-759) x US-09-345-882-4 (1-6002)

QY 40 LysTyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsn 59
 DB 1648 GAATTTGAGTGGATTGCGCAGAGAAAGTTGTTACAGCAATGTAAGGAGTGTGAAAT 1707
 QY 60 LeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysLysLysLysLys 79
 DB 1708 GTAAAGAA-----ATAAAGTTAAGGAGGAAATGAA 1740
 QY 80 GlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluLeuArgHis 99
 DB 1741 ACAGAGATCAAGAAATAAAGATGGAGGAGGAGGAAT-----ATAATACCA 1788
 QY 100 ArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAsp 119
 DB 1789 AGCAAGAAAGCCTATTGAGGAT-----GAAATTGAA 1821
 QY 120 GlnGluGlnGlnHisPro-----ArgSerArgLeuArgSerProPro 134
 DB 1822 AGAAAGAAATAATTAGCCCTCTCTGGGAAGTAAAGAAAGATTTATTAGATCTATACCT 1881
 QY 135 GluAlaLeuValGlnGlyArgTyrProHisIleLys-----AspGlyGluAspLeu 151
 DB 1882 ACACATCTCTGATCAGGAAAGAAAGTAAACATTAATAAAACCAAGACATGAAATCTG 1941
 QY 152 LysAspHisSerThrGluSerLysLysMetGluAsnCysLeu----- 165
 DB 1942 GATGCAAGATGATGACACAACTAGGTTAGTCAATCCCTCAACATAAAGTAGAAGCT 2001
 QY 165 ----- 165
 DB 2002 GAGGAGAAAGAAATCTCGAGATGAACGAATAAAGAAAGATGAAGATGATGAA 2061
 QY 166 ---GlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAspAlaSer 184
 DB 2062 GAAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2121
 QY 185 GlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGluLysGly 204
 DB 2122 AATGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2181
 QY 205 GluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer----- 220
 DB 2182 CGAGGAAAGAAATCAAAATAATGTAAGGCTAGTATTAAAGATTTCTGATGTTGAAGTGA 2241
 QY 221 -----GlyA-glyLysIleSerGluAsnSerTyrSerLeu 231
 DB 2242 GAGGTCCTTTACTTGTGTCATTCTCGGATGGAATGTGAGATACGATGAATGGATTAA 2301
 QY 232 AspAspLeuGluIleGlyProGlyGlnLeuSer----- 242
 DB 2302 GCAGATAAAATAGTAAGACCTCTCATATAAATATGTCACAAAGATAAATACATCGAGAAA 2361
 QY 243 ---SerSerThrPheAspSerGluLysAsnGlu-----SerArgArgAsnLeu 257
 DB 2362 ATAAAGATAAATATAGCAAAAGAAAGAAAGCAAAAGATGAAAAATACTCTCCAAAAAATGT 2421
 QY 258 GluLeuProArgLeuSerGluThrSerIleLys-----AspArgMetAla 272
 DB 2422 AAACITCGGCGCTGTGCAAAACCACCATTTTCAGACAAATCCATCTCTCGAAATGGTATCC 2481
 QY 273 LysTyrGlnAlaAlaValSerLysGlnSerSer----- 284
 DB 2482 AAACGTGATCTCACTGATGCCAAACAACTCTGTACTGTCTATTAAGTCCATAGAAATT 2541

QY 285 ThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGlu 304
 DB 2542 ACTTCGATCTTAATGGACTTCAAGCTTCTGAAGTTCGTGCAAGACAGTGCAGCAGAA 2601
 QY 305 GlnLysGluAsnValProProGlyProGluValCysIleThrHisGlnGluGlyGluLys 324
 DB 2602 GATGAGAGAGGTCTCAACACATGATATATATGCAAGAGAAATCTAAGATTGATCAT 2661
 QY 325 IleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArg 344
 DB 2662 TTGACCAACAACAGAAATGATCTTATT-----TCAAGGAGGAGAAACAGAAC 2706
 QY 345 AspSerGlnValLysSerGlu-----ValGlnGlnProVal 356
 DB 2707 AGTTCATCTTCTTAGAAGAAACAAAGTTCATGAGATTTTGGTAAATATCCAAACACG 2766
 QY 357 HisProLysPro-----LeuSerProAspSer----- 365
 DB 2767 TCAAAATCTCCAGAAAGATTAAAGAAAGATATAGAAGTATTATCCGAAGATCTGATTAT 2826
 QY 366 -----ArgAlaSerSerLeuSerGluSer 373
 DB 2827 GAAGAAGATGAAGTCACAAAGAGAGAGGATGTCAAGAGGACACAAACAGATAAATCT 2886
 QY 374 SerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCys-----Val 391
 DB 2887 TCAAAACCA-----CAAATAAACCTGTGTAAGAGAGGTATTGCAATACAGAA 2934
 QY 392 GluCysGlnLysThrValTyrPro-----MetGluArgLeuLeuAlaAsnGlnVal 409
 DB 2935 GAGTGTCTAAAACTGGATCACCTGGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2991
 QY 410 PheHisIleSerCysPheArgCysSerTyrCys----- 420
 DB 2992 -----TCATTTGCTATGGAACACAGTAGCAACAGCTCTTTCA 3027
 QY 420 ----- 420
 DB 3028 GATGAAGATGAAGAGAAACAAAGAGATGACACCACTAAGAAATAACAATGGTTG 3087
 QY 421 ---AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
 DB 3088 GAGGAAAAAGAAATCTCTACGACCACTGGTTTCTATTACGGA----- 3132
 QY 440 ProHisPheAsnGlnLeu-----PheLysSerLysGlyAsnTyrAspGluGly 455
 DB 3133 -----TTTTCAGAGTGGCGAGAAAAAGGATTAACCTTTTAAATAACTCTGTATGAAGA 3186
 QY 456 PheGlyHis-----ArgProHisLysAspLeuTrpAlaSer----- 467
 DB 3187 CTTCAAAACAGCAGCGGCCAAAGATCGAAAAAGATCTCTGGTCAAGTATTCAGGGACAGTGG 3246
 QY 468 ---LysAsnGluAsnGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
 DB 3247 CCTAAAAAAGCGTGAAGAGCTTTTTTCA--GACTCTGATCTAGGCTGAGCTTCC 3303
 QY 487 ThrProHisSerProGlyValGluAspAla-----ProIleAla 499
 DB 3304 CCACCGCAT---CCTGCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360
 QY 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
 DB 3361 GAAGAGAGAGTGTGTCACCCAGTGTAGAACTAGAAAAACCACTCCAGTCAATGTGAT 3420
 QY 520 AspLysProAlaGluThrLysLysLeuArgIle-----AlaTrpPro--- 533
 DB 3421 AGTAACCCATTGAAGAAAAACAGTAGAGGTCAATGACAGAAAAACAGAAATTTCCAAGT 3480
 QY 534 -----ProProThr-----GlyLeuGlySerSer--- 541
 DB 3481 AGTGGCAGTAATTCAGTGTGTAAATACCCCTCTCTACTACACCTGAATCGGCTTCATCAGTC 3540
 QY 542 -----GlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpPro 557

```
Db 3541 ACTGTAACAGAGGCGCGGAGCTCTCTGTAACAGATATCAGAACCACTGGCTCCA 3600
Qy 558 ProGluAspGluIleSer-----LysProGluValProGluAspValAspLeuAsp--- 574
Db 3601 AACCAAGAAGAGGTTCCGAAGTATCAAGAGTGAAGTATAGCACAATTCAGGTGGATAGT 3660
Qy 575 -----LeuLysLysLeuArgSerSerSerSerSerSerSerSerSerSerSerSerSerSer 590
Db 3661 GTTGTGGGGAGCTCCAGACCTCCAGTCTGAAGGGAAT-----AGTCCGCCA 3708
Qy 591 PheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerPro 610
Db 3709 GCAGGTTTGTGTCAGCTGAGCTCAAGCAGTATAGTATAGCAGCAGCAGCAGCAGCAGT 3768
Qy 611 ProIleArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGlyArgVal 630
Db 3769 -----GAAAGAGCTGTACAGGTGAGAAAGAGTGAAGATGCTCAGGAGGAGGAAGT 3822
Qy 631 AlaGluArgLysGlnValGluAsnAlaLysAla-----SerLysLysAsnGly 646
Db 3823 TCATCAAAAAGCAGAAAGAGCCATAAGCAACAGTGGTAAACACAAAAGAGGGA 3882
Qy 647 AsnValGlyLysThrTyrPheAsnLysGluSerLysGlyGluThrGlyLysArgSer 666
Db 3883 AAAGGCACAAATAGTAGTATAGTGAAGACTTTCATAGTACCAAAATCTCCGCGAAG 4002
Qy 667 Lys-----GluGlyHisSerLeuGlu-----MetGlu 675
Db 3943 CAGCCAGTCAAAATCAGTTCCAGTGAATGAAGTCTCATAGTACCAAAATCTCCGCGAAG 4002
Qy 676 AsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLys 695
Db 4003 ACGCAGTCTCCAGGAAATGTGGAAGATGGTATAGGATCTGTAT-----CTCAAG 4056
Qy 696 GlnGlnSerProGlnGluProLysSerLeuAsnTyrPheSer-----SerPheValAsp 712
Db 4057 GAACCCAGTAAATCGATTACCCAAAGTTTACAAATGGAGTTTTCAGATGTCGACCTGGAA 4116
Qy 713 Asn---ThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeu 731
Db 4117 AATATGACAGTGGCGAAGCATCACAATTTTCAAGAAAAATTTCAAGAAATCAGAAAA 4176
Qy 732 TrpGluGlyGluValValLysGluLeu---SerValGluGluGlnIleLysArg 748
Db 4177 CATTATCTGTCATTAAATCTGAAGTAGCTTCCATTGTCGAGGAGAAAGCGT 4230

RESULT 10
US-09-282-146-1
; Sequence 1, Application US/09282146A
; Patent No. 6303847
; GENERAL INFORMATION:
; APPLICANT: KAWAKURA, Akiyoshi
; APPLICANT: EBINUMA, Hiroyasu
; TITLE OF INVENTION: TRANSCRIPTION FACTOR CONTROLLING PHENYLPROPANOIC
; FILE REFERENCE: 4859-0027-0
; CURRENT FILING DATE: 1999-03-31
; EARLIER FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 988
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (100)..(702)
; NAME/KEY: misc feature
; LOCATION: (127)..(282)
```

```
; OTHER INFORMATION: LIM domain
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (427)..(582)
; OTHER INFORMATION: LIM domain
US-09-282-146-1
```

```
Alignment Scores:
Pred. No.: 2,77e-11 Length: 988
Score: 212.50 Matches: 51
Percent Similarity: 46.58% Conservative: 24
Best Local Similarity: 31.68% Mismatches: 56
Query Match: 5.41% Indels: 30
DB: 4 Gaps: 4
```

US-09-890-549-4 (1-759) x US-09-282-146-1 (1-988)

```
Qy 384 AlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMetGluArgLeu 403
Db 109 GCAGGAACACACAGAAATGCGATGTCATGTCAGCAAGACTGTCTATCTGTTGACAATTA 168
Qy 404 LeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLys 423
Db 169 ACTGCAGATAACAGAAATCTATCACAAGAGCTGTTTCAGATGCCATCCTGCAAGGCACT 228
Qy 424 LeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsn 443
Db 229 GTCAAGCTTGGCACTACAAATCTCTTTGAGGAGTCTTATCTATGTAGACCACACTTTGAT 288
Qy 444 GlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAsp 463
Db 289 CAGCTCTTCAAAACAACTGGCAGTTTGGATAAAAGCTTTGAAGGTACACCA----- 339
Qy 464 LeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGlnLeuAlaAsn 483
Db 340 -----AAAAATGTGAAG----- 351
Qy 484 AlaArgGluThrProHisSerProGlyValGluAspAlaPro---IleAlaLysVal--- 501
Db 352 -----CCACAGAAACCCATTGACGTGAGAAACCAACAGGTAGCCAAAGTGACA 399
Qy 502 -----GlyValLeuAlaAlaSer-MetGluAlaLysAlaSerSerGlnGlnGln 517
Db 400 AGCATGTTTTGGTGGAAACAGAGAGAAATGTTTTGGCTGCAGAGAAACTGTCTATCCCAACA 459
Qy 517 uLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGln 537
Db 460 GAAAGAGGTATCAGCAATGGCAGCCATACCATAGAGCTGCTTCAATGCAGCCACCGA 519
Qy 537 u 537
Db 520 G 520
```

```
RESULT 11
US-08-934-627B-1
; Sequence 1, Application US/08934627B
; Patent No. 6169174
; GENERAL INFORMATION:
; APPLICANT: OSAMU HASEGAWA
; APPLICANT: SATOSHI AOTSUKA
; APPLICANT: SOICHIRO TAKENISHI
; APPLICANT: HIROFUMI UCHIMIYA
; TITLE OF INVENTION: COTTON PLANT GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 3424 Peachtree Road, N.E., 2400 Monarch Tower
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,627B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Roger T. Frost
REGISTRATION NUMBER: 22,176
REFERENCE/DOCKET NUMBER: 20111-0010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 949-2400
TELEFAX: (404) 949-2499
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum L.
FEATURE:
NAME/KEY: CDS
LOCATION: 134..757
US-08-934-627B-1

Alignment Scores:
Pred. No.: 1,4e-10 Length: 1015
Score: 205.50 Matches: 44
Percent Similarity: 52.00% Conservative: 21
Best Local Similarity: 35.20% Mismatches: 59
Query Match: 5 Indels: 1
DB: 1 Gaps: 1

US-09-890-549-4 (1-759) x US-08-934-627B-1 (1-1015)

QY 379 MetLysPheGlnAlaProLysGluThrCysValGluCysGlnLysThrValTyr 398
Db 134 ATGGCGACATTTCCAGGA---ACTCAACAAAATGCAATGCGATGCAACAGACTTTAT 190
QY 399 ProMetGluArgLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSer 418
Db 191 CTGGTGGAATGAGTACGGCTGATATAGGGTTTTTCATAGGCTTGCTTCGTTGCCAC 250
QY 419 TyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCys 438
Db 251 CATTGCAAGGTACCTCAAGCTTAGCAACTACAACTCATTCATTGAAGGGTGCTACTATGC 310
QY 439 LysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHis 458
Db 311 AGGCCACACTATGATCACTCTTCAAGAGAACTGCGACTCTTGACAGAGTTTGAAGGA 370
QY 459 ArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluLeuLeuGluArgPro 478
Db 371 ACACCAAGGTTGTCAAACTGAAAGACAAATCGATAGTGAAGTGCCTCAAGAGTATG 430
QY 479 AlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIle 498
Db 431 AACTCGTTGGTGGCACCAGAGAAAATGTGGCGCTGTCTTAAGACTGCTTATCCAATT 490
QY 499 AlalysValGlyVal 503
Db 491 GAGAGGGTTACTGTG 505

RESULT 12

US-09-620-312D-134
Sequence 134, Application US/09620312D
Patent No. 6569662

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-bong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 134
LENGTH: 4323
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (81)..(3920)
US-09-620-312D-134

Alignment Scores:

Pred. No.: 1,34e-08 Length: 4323
Score: 196.00 Matches: 154
Percent Similarity: 33.49% Conservative: 125
Best Local Similarity: 18.49% Mismatches: 324
Query Match: 4 Indels: 230
DB: 4 Gaps: 35

US-09-890-549-4 (1-759) x US-09-620-312D-134 (1-4323)

QY 13 SerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysAsnLysSerSer 32
Db 636 AACTTAATGTGCAAGCAAGCAAGCCCTTGAAGAAAGCTACA---GGGAAGGATTC 692
QY 33 AlaIleValGluIlePheSerLysTyrGlnLysAlaAlaGluThrAsnMetGluLys 52
Db 693 CCATGCTCAGCATCGCAGGACCCCTCCAGCAGATCCACA---GATTCAGTATGGAAATC 749
QY 53 LysArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeu 72
Db 750 TCAATTTC---ACTGAATGCTGGATGAACCCGAAAGAAATGGATCAACAGAGAGAA 806
QY 73 LysLysLysLys-----TyrGluAsnProGlyLeuGlyAlaGluSerHis 86
Db 807 AACCTTCCTGGACATAACTATCTTTTGTGATTTCTCCAGGACAGATGATGAAAAACAGT 866
QY 87 ThrAspSerLeuArgAsnSerSerThrGluIleArgHisArgAlaAspHisProAla 106
Db 867 GTAGATAGTGTCAAAATTCCTACTGTAGCCATAAATCTCGA-----CCTGTTTCA 917
QY 107 GluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGlnGluGlnIleHisPro 126
Db 918 AGAGTTACCAATGGAACCTTCCAAT-----AAAAAAGTATTATGNA 959
QY 127 ArgSerArgLeuArgSerProGluAlaLeuValGlnGlyArgTyrProHisIleLys 146

Db 959 ----- 959
Qy 147 AspGlyGluAspLeuLysHisSerThrGluSerLysLysMet-----GluAsnCys 164
Db 960 -----CAAGACACTAATAATAACAGTACTAATAACAGAAAGTCAAGTGGCAAGGATGT 1013
Qy 165 LeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluLysSerGluLysSerGluLysSer 184
Db 1014 AGTGAGCCAGTACCACAGGCAATTTTGAAGAAAGAGGAACTAGCAATGGATGTACTGCA 1073
Qy 185 GlyLysIleGluLysArgAsnValProLeuAsnArgLeuLysMetMetPheGluLysGly 204
Db 1074 GCTCAGCAGACAGCAAGAGTAGTACCCCACTAATCTTACTATAAAGTCAAGGATCCCAAGGA 1133
Qy 205 Glu-----ProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSerGlyArgLys 223
Db 1134 GAGTCACCAAACTCAGTAAATA-----TCTTCAGTCTCTCAAGGCGAG 1175
Qy 224 IleSerGluAsnSerTyrSerLeuAspAspLeuGluLysGlyProGlyGlnLeuSerSer 243
Db 1176 TCTGATGAAATGTGGCAAGTTGGAC----- 1202
Qy 244 SerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuProArgLeuSer 263
Db 1203 CACAATACAACTACAGAGAAACACACCTAAGAGAAATG----- 1244
Qy 264 GluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLysGlnSerSer 283
Db 1245 GTCAAGCAAGTACACACAGCTTTGGCTAAGGTTAATGCAAAATAGTGGCAATCCCTAAA 1304
Qy 284 SerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluLysIleHisLysMet 303
Db 1305 AATCTAATTCAGTCAAAATA-----GGTGAA-----ACTTTGAATAATAA 1346
Qy 304 GluGlnLysGluAsnValProGlyProGluValCysIleThrHisGlnGluGlyGlu 323
Db 1347 GATTCAAAACAGAAATGCTCTCGACAGGTATATCAAAATCAGCCCTTCCTCCCAA 1406
Qy 324 LysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGluAspSer 343
Db 1407 AGACCTTTAAACATGAAACATCTACTGTCCAAAAAGT-----ATGTTTCATGATGTG 1460
Qy 344 ArgAspSerGlnValLysSerGluValGlnProValHisProLysProLeu----- 361
Db 1461 CGTGATAATAACAACAGGACAGTGTCTTCTGAACAGAGGCTCACAACCTCTCATTAAT 1520
Qy 362 -----SerProAspSerArgAlaSerSerLeuSerGluSerProLysAlaMet 379
Db 1521 CTTCATCTGAAATAAGTATGACAGAGCACTCCAGTCATCTCTSCAGGCGCTGACCCACAA 1580
Qy 380 Lys----- 380
Db 1581 AAGCCATTAAACGATCAAGAAAAAGAGAGTGGCGTTAGATGCCAAAAATATTTCAAAG 1640
Qy 381 -----LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1641 CTGGATAAATCAATTAACACAGCACTGGAATCAAAACAGATTTGTTTAGATAAAGTGAA 1700
Qy 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPhe 415
Db 1701 ACAAAATTTCCCAATCAAAAGAAACAGATGATGGCGATGCAGCTAACATATGTTGTCTAT 1760
Qy 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1761 -----TCTGTGGGAGT---GATATGTAAATTCANAA 1790
Qy 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1791 TTTTATAGCACCACAGCCCTAAAAATACATGGTTTCAATCCAAAT-----GAAACTCC 1844
Qy 456 PheGlyHisArgProHisLysAspLeu-----TTPAlaSerLysAsnGluGluLys 474

Db 1845 TTGAACCTCTTAATCCAGTTTGTGATTTAGACTCAACAGTGCAGGCAAAATCCATTGTGATA 1904
Qy 475 LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSer-----ProGly 492
Db 1905 TCAGATAGGAGACCAAGTAGGAGAAAGATACAAACAAACAATCAAGTATTAATGT 1964
Qy 493 ValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAla 512
Db 1965 GTGAAGATGTTTCACTGTGTAATCTCTGAAGGACAAATGGTACCTTAAT----- 2015
Qy 513 SerSerGlnGlnGluLysGluAspLys-----ProAlaGluThrLysLysLysLeuArgIleAla 531
Db 2016 TCTGCTCAAGAACACAAAAATCGAAAGTTCTCTGTGGAAGGAGCTGACAAT----- 2066
Qy 532 TrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysMet 551
Db 2067 -----CCTAGTAGTTG---TCAGATGAATCTGCTATGGATGAA----- 2102
Qy 552 SerLysProLysTrpProProGluAspGluIleSerLysProGluValProGluAspVal 571
Db 2103 -----GACAAACATGCTCAGCA 2120
Qy 572 AspLeuAspLeuLysLysLysLeuArgArgSerSerLeuLysGluArgSerArgProPhe 591
Db 2121 GACTCAGATGATCTTCCAAAGTGTCTTTCGGACAGCTATCAGAAAAAAATCTCCTATAA 2180
Qy 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
Db 2181 AATATGGAACATCAGAACTCTCCAGAGACCCATGAACACTCCAGAACT-----CCA 2231
Qy 612 IleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGlyArgValAla 631
Db 2232 TTTGTGGTCACTGGATTTGAGTACT-----GGTGTCTGCGAT 2270
Qy 632 GluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThr 651
Db 2271 CAGCGAGAGAGTCTGAAATCT----- 2291
Qy 652 ThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSer 671
Db 2292 -----GACACTGGCAGTGTACCACCTCTCCCGATGAC 2124
Qy 672 LeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsn 691
Db 2325 ATAAAGCCAGATCTGAAGACTATGATGCTGGAGGCTCTCAGGATGATGGTCAAT 2384
Qy 692 -----SerPheLeuLysGlnGlnSer 698
Db 2385 GACAGAGGTATCTTAATGTGGCACTATGCTGCCATGATTTTCTGGAAAGAGTAGC 2444
Qy 699 -----ProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsn 713
Db 2445 AGTGATACCAGTACTCTGAGAAATTAATA-----ATTATGATAGT 2486
Qy 714 ThrPheAlaGluLupPheThrThrGlnAsnGlnLysSerGlnAspVal----- 729
Db 2487 AATTTAAGAAATGAAGTAAAAATGAAAGCAAGTAAATGATGATCTTTTCCAGTTAAT 2546
Qy 730 -----GluLeuTrpGluGlyGluValVal 737
Db 2547 TCAACGAGTGTATGATAATCTTAGAAAAAGGCCAGAAATTTGGTCTCGATCTGCAATA 2606
Qy 738 LysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyr----- 752
Db 2607 GTTCACTCTAGGAAAGAGAAATATTCACAGAGGAGTGTCCAGTTTGGCTCAGGAAATA 2666
Qy 753 -----AspGluAspGluAspGlu 758
Db 2667 GATCAGGTATCTTCTTCAGCAGATGAAACAGACATGAA 2705

RESULT 13
US-09-220-132-79
; Sequence 79, Application US/09220132

```

; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 5857
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-220-132-79

```

Alignment Scores:	
Pred. No.:	2.49e-09
Score:	195.50
Length:	5957
Matches:	163
Conservative:	18
Mismatches:	304
Indels:	25
Gaps:	3
DB:	4
Query Match:	4.98%
Best Local Similarity:	35.49%
Percent Similarity:	19.88%

US-09-890-549-4 (1-759) x US-09-220-132-79 (1-5857)

QY	7	AsnArgGlnTrpThrSerLeuSerLeuArgValThrAlaLysGluLeuSerLeuVal	26
Dd	1795	CACCAGAGAAATAACTTCTCGAAGGAGCATTTTGAGCCGGGAAGAACAAC-	1848
QY	27	AsnLysAsnLysSerSerAlaIleValGluIlePheSerLysTyrGlnLysAlaAlaGlu	46
Ddb	1849	CATCAGAAGAGATAAAGGCTGTATACGCCACCGAAAAGCTTCCAAAGACAACGAG	1908
QY	47	GluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeuSerGlnHisPheArgLys	66
Ddb	1909	TCATTGAAAAAGCAAGCTGGAGATGCCAACAAAGAGAAC-----TCA	1950
QY	67	GlyThrLeuThrValLeuLysLysLysTrpGlu-----	77
Dd	1951	GATCTGATAGCTCTATGGAACTCCAAACTGGAGACTGCCATCCCATCCACGAGCGC	2010
QY	78	-----AsnProGlyLeuGlyAlaGluSerHisThrAspSer	89
Dd	2011	ATGGAAGAACTGAAGGTATCTTTCAGCAAAGGGCTTGCACAGAG-----ACGGCAGAA	2064
QY	90	LeuArgAsnSerSerThrGluIle---ArgHisArgAlaAspHisProProAlaGluVal	108
Dd	2065	TTTGCTGAACATAAAACACAAATAGAGAAAATGAGACTAGATTAC-----	2109
QY	109	ThrSerHisAlaAlaserGlyAlalysAlaaspGlnGluGlnIleHisProArgSer	128
Dd	2110	---CAACACGAATAGAAAAATTGCGAGAATCAACAAGACTCTGAA-----	2151
QY	129	ArgLeuArgSerProProGluAlaLeuValGlnGlyArgTyrProHisIleLysaspGly	148
b	2152	-----CGGCTGCCCATGCTCAAGAGATG	2175
Y	149	GluAspLeuLysAspHisSerThrGlu---SerLysLysMetGluAsnCysLeuGlyGlu	167
b	2176	GAAGCCTTGAGGGCTTAACACTGATGAAGTATTATAAGAAAAAGGAAAACAGTCTGNAGCC	2235
Y	168	SerArgHisGluValGluLysSerGlu-----IleSerGluAsnThrAspAlaSer	184
b	2236	ATCAGTGCMAACTGGCAAGCAGAGAACCCAGCATCTCGTAGAAATGGAAGACACGTTA	2295
Y	185	GlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGluLysGly	204
b	2296	AACAAATTACAGGAAGCTGAATAAAGCTAAAGCAGCTAGAGTACTGCAAGCCAATGC	2355

```
QY 564 LysProGluValProGluAspValAspLeu-----AspLeuLysLysLeu 578
Db 3082 AAAAAAGCATGAGGAGAAAGAAAGAAATTTGAGAGGAAATTTGCGACCTGGAAAAAGAAA 3141
QY 579 ArgArgSerSer-----SerLeuLysGluArgSerArgProPheThr--- 592
Db 3142 ATGGAACCAAGCCACACACCACTGTGAGAGCTGAAAGCCAGGTATGAGAGAGCCACTTCT 3201
QY 593 -----ValAlaAlaSerPheGlnSerThrSerValLysSer 604
Db 3202 GAGACAAAAACAAGCATGAGAAATCTCAGAAACCTCCAGAAAGCGCTGCTGACACA 3261
QY 605 ProLysThrValSerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGlu 624
Db 3262 GAGGACAAAGTGAAGGGCGCAGGAGGAGCAAGTGTCTGTCGAGGAGCTGGAGGAG 3321
QY 625 SerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLys 644
Db 3322 CTG-----AGAAAGCAAGCCAGCAAGCAAGCAAGCAAGCTGCTCAACA 3360
QY 645 AsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGlyGluThr----- 662
Db 3361 GCGGAAGATGCCATGAGATGAGAAAGTGAAGAGCTGCAACAAATCAAAAGCAAGTGA 3420
QY 663 -----GlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 3421 TCCTTGAGGACACCAAGCAAAATGCAAAATGCAAAATGCAAAATGCAAAATGCAAAAT 3480
QY 581 GluAsnGlyAlaAspSerAspGluAspAsn----- 691
Db 3481 GAAACAACTTGAAATGTGAAGAGCTGCAACAAATGCAAAATGCAAAATGCAAAATGCA 3540
QY 692 -----SerPheLeuLysGlnSerProGln 700
Db 3541 AATCAAAAATGGAAGATTTAGAAAGAAATAGAAACCTTAAGCAGCAGCAGCTCAG 3600
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe--- 719
Db 3601 AAGTCCAGCAGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAG 3660
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGlu 739
Db 3661 AGAAGCAGGAGCAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3720
QY 740 LeuSerValGluGluGlnLysArgAsnArgTyTrpAspGluAspGluAspGluGlu 759
Db 3721 AATCAGTTGTTGAAATGAAAGAAAGAGAAATCAAGTTCATAAAGACGACATGAAGAG 3780

RESULT 14
US-08-452-655B-1
; Sequence 1, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF INVENTION: 102
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESS: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP2.5(APC)
US-08-452-655B-1
```

```
Alignment Scores:
Pred. No.: 6,54e-08 Length: 8532
Score: 194.00 Matches: 153
Percent Similarity: 35.44% Conservative: 127
Best Local Similarity: 19.37% Mismatches: 289
Query Match: 4.94% Indels: 221
DB: 1 Gaps: 38
```

US-09-890-549-4 (1-759) X US-08-452-655B-1 (1-8532)

```
QY 26 ValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLysLysGlnLysAlaAla 45
Db 5365 GTAAGAAAAAATGCAGAC-----TCAAAAAATAATTTAAATGCT 5403
QY 46 GluGluThr-----AsnMetGluLysLysArgSerAsnThrGluAsnLeuSerGln 62
Db 5404 GAGAGAGTTTCTCAGACAAACAAAGATTCAAGAAACAGAAATTTGAAAAATAATTCACAG 5463
QY 63 HisPheArg----- 65
Db 5464 GACTTCATGATGATGCTCCCAAAATATAGATAGATGATGATGATGATGATGATGATGATGAT 5523
QY 66 -----LysGlyThrLeuThrValLeuLysLysLysLysLysLysLysLysLysLys 77
Db 5524 TCACCTCATCATACAGCGCTATTGAGGAACTCTTACTGTTTTCACGAAATGATTTCT 5583
QY 78 AsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIle 97
Db 5584 TTGAGTTCTCTAGATTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5643
QY 98 ArgHisArgAlaAspHisProAlaGlu-----ValThrSerHis-----Ala 112
Db 5644 AGAAGCGCAAAAGAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAG 5703
QY 113 AlaSerGlyAlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSer 132
Db 5704 TCCAAACCAATCAGCTAATAAGACACAAAGCTATTGAAAGACCAATTAATTCAGGCT 5763
```


Search completed: January 7, 2004, 01:47:34
Job time : 153 secs

```
QY 395 LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCys 414
Db 6538 AAG----- 6540
QY 415 PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly 434
Db 6540 ----- 6540
QY 435 ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu 454
Db 6541 -----ATAGAAATCTGAAAGTAAAGGAATCAAA 6567
QY 455 GlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys---AsnGluAsnGluGlu 473
Db 6568 GGA---GGAAAAAAGATTATATAAGTTTGATTACTGGAAAAAGTTCGATCTAATTCAGAA 6624
QY 474 Ile-----LeuGluArgProIleGln-----LeuAlaAsnAlaArg 485
Db 6625 ATTTCCAGCCCAAAATGAAACAGCCCTTCAAGCAAAACATGCCTTCAATCTCTCGAGGCAGG 6684
QY 486 GluThrProHisSerProGlyValGluAsp-----AlaProIleAlaLys 500
Db 6685 ACAATGATTCATATTTCCAGGAGTTCGAAATAGCTCTCAAGTACAAAGTCCTGTTCTTAAA 6744
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGlu----- 517
Db 6745 AAAGCCCAACCCCTTAAGACTCCAGCCTCCAAAAGCCCTAGTGAAGTCAAAACAGCCCAAC 6804
QY 518 -----LysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProPro 534
Db 6805 ACTTCTCTAGAGGAGCCAGCCATCTGTGAATCAGAAATTAAGCCCTGTGTCAGGCAG 6864
QY 535 ProThrGluLeuGlySerSerGlySerAla---LeuGluGluGlyIleLysMetSerLys 553
Db 6865 ACATCCCAATAGGTGGTCAAGTAAAGCACCTTCTAGATCAGGATCTAGAGATTCGACC 6924
QY 554 ProLysTyrProProGluAspGluIleSerLysProGluValProGluAspValAspLeu 573
Db 6925 CCTCAAGACCTGCCAGCAACCATTAAGTAGACCT----- 6960
QY 574 AspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSer-----ArgProPhe 591
Db 6961 ---ATACAGTCTCTCGCCGAACTCAATTTCCCTCGTAGAATGGAATAAGTCTCTCT 7017
QY 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
Db 7018 AACAAATTTATCTCAACTTCCAAGGACATCA-----TCCCTAGTACTGCTTCA----- 7065
QY 612 IleArgLysGlyTyrPsrMetSerGluGlnSerGluGluSerValGlyGlyArgValAla 631
Db 7066 ---ACTAAGTCTCAGGTTCTGGAAAAATGTCTATATACATCTCCAGGTAGACAGATGAGC 7122
QY 632 GluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlnLysValGlyLysThr 651
Db 7123 CAACAGAACCTTACCAACAAACAGGTTTATCCAGAAATCCAGTAGTATTCCAGAAAGT 7182
QY 652 ThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSer 671
Db 7183 -----GAGTCTGCCTCCAAAGGACTAAATCAGATGAATATGTAATGGAGCCAAAT 7233
QY 672 LeuGluMetGluAsnGluAsnLeu-----ValGluAsnGlyAlaAspSerAspGlu 688
Db 7234 AAAAAGGTAGAACTTTCTAGAATGTCTTCAACTAAATCAAGTGAAGTGAATCTGATAGA 7293
QY 689 AspAspAsnSerPheLeuLysGlnGlnSer-----ProGln-GluPr 702
Db 7294 TCAGAAAGACCTGTATTAGTACCCAGTCACTTTTCATCAAGAGAGCTCCAAAGCCCAAC 7353
QY 702 cLysSerLeuAsnTrpSerPheVal 711
Db 7354 TTAAGAGAAATTTGGAGGAATCTGCTT 7381
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:29:06 ; Search time 606 Seconds
(without alignments)
4335.162 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSPFNRROWTSLSLRVTA.....LSVEEQIKRNYVDEDEDEE 759

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2_1/USPTO_spool/US09890549/runat_06012004_094753_19859/app_query.fasta_1.903

-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09890549@cgn_1_1_232 @runat_06012004_094753_19859

-NCPU=6 -ICPU=3 -NO_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
---------------	-------	-----------------------	-------	-------------

RESULT 1

US-10-117-722-45

; Sequence 45, Application US/10117722

; Publication No. US20030219744A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and

; FILE REFERENCE: 784CIP2BCIP

; CURRENT APPLICATION NUMBER: US/10/117,722

; CURRENT FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: 09/620,312

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1104

; SOFTWARE: pt_FL_genes Version 1.0

; SEQ ID NO 45

ALIGNMENTS

	1	3927	100.0	2905	13	US-10-117-722-45	Sequence 45, Appl
	2	3927	100.0	2905	15	US-10-037-270-45	Sequence 45, Appl
	3	3872.5	98.6	3650	11	US-09-783-732-3	Sequence 3, Appl
	4	3603.5	91.8	3543	11	US-09-783-732-1	Sequence 1, Appl
	5	2535	64.6	2749	9	US-09-822-849A-399	Sequence 399, App
	6	1254	31.9	732	9	US-09-925-297-152	Sequence 152, App
	7	841.5	21.4	688	15	US-10-198-846-10895	Sequence 10895, A
	8	828	21.1	585	15	US-10-060-036-4338	Sequence 4338, Ap
	9	608.5	15.5	2710	13	US-10-094-749-1019	Sequence 1019, Ap
	10	608	15.5	3236	10	US-09-833-381-1727	Sequence 1727, Ap
	11	596.5	15.2	3825	11	US-09-909-567B-2	Sequence 2, Appl
	12	546.5	13.9	2955	12	US-10-161-927-73	Sequence 73, Appl
	13	534.5	13.6	2379	9	US-09-880-192-13	Sequence 13, Appl
	14	534.5	13.6	2379	13	US-10-427-348-13	Sequence 13, Appl
	15	385	9.8	1298	13	US-10-440-366-1	Sequence 1, Appl
	16	383	9.8	3178	12	US-10-104-047-989	Sequence 989, App
	17	381	9.7	705	10	US-09-833-381-1735	Sequence 1735, Ap
	18	370.5	9.4	1299	10	US-09-833-381-1733	Sequence 1733, Ap
	19	367	9.3	1081	9	US-09-789-919-5	Sequence 5, Appl
	20	363	9.2	641	10	US-09-833-381-1730	Sequence 1730, Ap
	21	331.5	8.4	397	9	US-09-789-919-41	Sequence 41, Appl
	22	327	8.3	523	10	US-09-833-381-1736	Sequence 1736, Ap
	23	290	7.4	228	10	US-09-783-580-176	Sequence 176, App
	24	277	7.1	494	11	US-09-918-995-14789	Sequence 14789, A
	25	273.5	7.0	743	9	US-09-789-919-6	Sequence 6, Appl
	26	245	6.2	8977	12	US-10-359-012-5	Sequence 5, Appl
	27	240.5	6.1	6008	12	US-10-359-012-3	Sequence 3, Appl
	28	235	6.0	3290	15	US-10-084-817-286	Sequence 286, App
	29	231.5	5.9	3609	10	US-09-917-800A-1715	Sequence 1715, Ap
	30	226	5.8	4226	13	US-10-117-722-480	Sequence 480, App
	31	226	5.8	4226	15	US-10-037-270-480	Sequence 480, App
	32	224.5	5.7	861	15	US-10-278-173-83	Sequence 83, Appl
	33	223	5.7	6002	13	US-10-126-704-4	Sequence 4, Appl
	34	223	5.7	6002	15	US-10-071-179-4	Sequence 4, Appl
	35	222	5.7	4205	14	US-10-098-841-330	Sequence 330, App
	36	221	5.6	14172	12	US-10-359-012-7	Sequence 7, Appl
	37	216	5.5	693	9	US-09-770-149-291	Sequence 291, App
	38	215	5.5	760	15	US-10-295-403-115	Sequence 115, App
	39	214	5.4	5763	14	US-10-002-600-79	Sequence 79, Appl
	40	213.5	5.4	396	10	US-09-878-574-3019	Sequence 3019, Ap
	41	213.5	5.4	418	10	US-09-878-574-301	Sequence 30, Appl
	42	212.5	5.4	908	10	US-09-928-412-1	Sequence 1, Appl
	43	212.5	5.4	9009	12	US-10-359-012-9	Sequence 9, Appl
	44	212	5.4	5659	14	US-10-002-600-80	Sequence 80, Appl
	45	211	5.4	5373	12	US-10-369-493-25273	Sequence 25273, A

/		LENGTH: 2905	
/		TYPE: DNA	
/		ORGANISM: Homo sapiens	
/		FEATURE:	
/		NAME/KEY: CDS	
/		LOCATION: (173)..(2452)	
/		US-10-117-722-45	
Alignment Scores:			
Pred. No.:	0	Length:	2905
Score:	3927.00	Matches:	759
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0
US-09-890-549-4 (1-759) x US-10-117-722-45 (1-2905)			
QY	1	MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla	20
Db	173	ATGGAAATCATCTCCATTTAATAGACGGCNAATGACCTCCTCATATCATTTAGGGGTAAACAGCC	232
QY	21	LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys	40
Db	233	AAAGAACTTTCTTGTCAACAAGAACCAAGTCTCGGCTATTGTGGAATATTCTCCAAAG	292
QY	41	TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu	60
Db	293	TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAAAGAGAGAAGTAACACCGAAAAATCTC	352
QY	61	SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly	80
Db	353	TCCAGACCTTTAGAAAGGGGACCTTGACTGTGTTAAAGAAAGAGTGGGAAACCCAGGG	412
QY	81	LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg	100
Db	413	CTGGGAGCAGAGTCTCACACAGCTCTCTACGAAACACAGCAGCCTGAGATTAGGCACAGA	472
QY	101	AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln	120
Db	473	GCAGACCATCTCTCTGCTGAAGTGAACAGCCAGCGTCTCTGGAGCCAAAGCTGACCAA	532
QY	121	GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly	140
Db	533	GAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAGGCCCTCGTTCAAGGT	592
QY	141	ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys	160
Db	593	CGATATCCCAACATCAAGACGGTGAGGATCTTTAAAGACCACCTCAACAGAAAGTAAAAAA	652
QY	161	MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn	180
Db	653	ATGGAAATTTGTAGGAGAAATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAANAAC	712
QY	181	ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet	200
Db	713	ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAACATGATG	772
QY	201	PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer	220
Db	773	TTTGAAAGGTGAACCAACTCAACTAAGATTCTTCGGGGCCCAAGCCGGAAGTGCAGT	832
QY	221	GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln	240
Db	833	GGAAAGAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCCCAGGTCAG	892
QY	241	LeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro	260
Db	893	TTGTCATCTTACATTTGACTCGGAGAAAAATCAGAGTAGACCAAGTAATCTGGAACTTCCA	952
QY	261	ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys	280
Db	953	CGCTCTCAGAAACCTCTATTAAGAGTTCGAATGGCCCAAGTACCAGGCGAGCTGTGTCCAAA	1012

QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
Db 2093 GCTTCTAAGAAGAAATGGGAATGTGGGAAACCAACCTGGCAAAACCAAGAACTTAAGGA 2152
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 GAGACAGGAGAGAGAGTAAGGAGGTATAGTTGGAGATGGAGAAATCTTGTA 2212
QY 681 GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGGTGAGACTCCGATGAAGATGAATACAGCTTCCTCAACCAACCAATCTCCCAA 2272
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCAAGTCTCTGAATGGTTCAGTTTGTAGACCAACACCTTTGCTGAAGAAATCACT 2332
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGluValValLysGluLeu 740
Db 2333 ACTCAGAATCAGAAATCCAGGATGTGGAATCTCGGAGGAGAGAGTGGTCAAGAGCTC 2392
QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2393 TCTGTGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAAGAG 2449

RESULT 2

US-10-037-270-45
; Sequence 45, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 45 -
; LENGTH: 2905
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (173)..(2452)
US-10-037-270-45

Alignment Scores:
Pred. No.: 0 Length: 2905
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-890-549-4 (1-759) x US-10-037-270-45 (1-2905)
QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 173 ATGGAATCATCTCCATTTTAAATAGCGCAATGACCTCACTATCATTTGAGGGTAAACAGCC 232
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 233 AAAGAACTTTCTTGTCAACCAAGAACCAAGTCAATCGGCTATTTGTGAAATATTTCTCAAG 292
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysLysLysLysLysLysLysLys 60
Db 293 TACCAAGAAAGCAGCTGAAGAAACCAACATGGAGAGAGAGAGAGTAAACACCAAAATCTC 352
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysLysLysLysLysLys 80
Db 353 TCCAGCACTTTAGAAAGGGGACCTGACTGTGTTTAAAGAGAGAGTGGAGAACCCAGGG 412
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 413 CTGGAGCAGAGACTCTCACAGACTCTCTACGGAACAGCAGCAGCTGAGATTAGGCACAGA 472
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 473 GCAGACCATCTCTCTGCTGAGTGAAGTGAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAA 532
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 533 GAAGAAACAAATCCACCCAGATCTAGACTCAGGTCCCTCTGAAAGCCCTCGTTGAGGT 592
QY 141 ArgTyrProHisIleLysAspGlyGluAspLysLysAspHisSerThrGluSerLysLys 160
Db 593 CGATATCCCAACATCAAGGACGCTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 652
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 653 ATGGAAATTTGCTAGGAGATCCAGGCATCAAGTAGAAAAATCAGAAATCAGTGAAGAAC 712
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 713 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAAGATGATG 772
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 773 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGTCAAGT 832
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
Db 833 GGAAGGAAGATCTCTGAAAACAGCTATTCTAGATGACCTGGAAATAGGCCCCAGGTGAG 892
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuLeuPro 260
Db 893 TTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAATCTGGAATCTCCA 952
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 953 CGCTCTCAGAAACCTCTATAAGAGATGGAATGGCCAGTAGTACCAGGAGCTGTGTCCAAA 1012
QY 281 GlnSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluIleLysIle 300
Db 1013 CAAGCAGCTCAACCACTATACAAATGAGCTGAAGCCAGTGTGGGAAATCAAAAT 1072
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 1073 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCACCCTCAG 1132
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1133 GAAGGGGAAAGATTTCTGCAAAATGAGATAGCTGCGAGTCCGTTCCACCCCTGCCGAA 1192
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360

Db 1193 GATGACTCCCGTGACTCCAGGTTAAAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCA 1252
Qy 361 LeuSerProAspSerArgAlaSerSerLeuSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1253 CTAAGTCCAGATTCACAGAGCTCCAGTCTTCTGAAAGTTCTCTCCCAAGCAATGAAG 1312
Qy 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1313 AAGTTTCAGGCACCTGCCAGAGAGACCTCGGTGGAAATGTCAAGAACAGCTCTATCCAATG 1372
Qy 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1373 GAGCGTCTCTTGCCCAACAGCAGAGTGTTCACATCAGCTGTTCCGTTGCTCTTATTCG 1432
Qy 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1433 AACCAACAACTCAGCTAGAACATATGCATCTTTACATGGAAGAATCTATTGTAAGCCT 1492
Qy 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1493 CACTTCATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACACAGCA 1552
Qy 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluLeuLeuGluArgProAlaGln 480
Db 1553 CACAAGGATCTATGGCAGAGCAAAATGAAACGAGAGATTTTGGAGAGACCCAGCCAG 1612
Qy 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1613 CTTGCAAAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTCTTAAG 1672
Qy 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 1673 GTGGGTGTCCTGGCTGCAAGTATGAAGCCAAAGGCTCTCTCAGCAGGAGGAAGAC 1732
Qy 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
Db 1733 AAGCAGAGTGAACCAAGAGCTGAGGATCGCTGGCCACCCCACTGAACTTGAAGT 1792
Qy 541 SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProProLysAsp 560
Db 1793 TCAGGAAGTGCTTGGAGGAGGATCAAAATGTCAAGCCCAATGATGCCCTCTGAGAC 1852
Qy 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db 1853 GAAATCAGCAAGCCGCAAGTCTCAGGATGTCGATCTAGATCTGAAGAAGTAAGACGA 1912
Qy 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1913 TCTTCTTCACTGAGGAAGAAGAGCCGCCATTCATCTGAGCAGCTTCAATTCAAAGCAC 1972
Qy 601 SerValLysSerProLysThrValSerProPheArgLysGlyTrpSerMetSerGlu 620
Db 1973 TCTGTCAAGAGCCCAAAACTGTGTCACCTATCAGGAAGGCTGGAGCATGTCCAG 2032
Qy 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 2033 CAGAGTGAAGAGTCTGTGGTGGAAAGTTGCAGAAAGGAAACAAGTGGAAAAATGCCAAG 2092
Qy 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLysGly 660
Db 2093 GCTTCTAAGAGATGGGAATGTGGAAAAACAACCTGGCAAAACAAGAAATCTTAAAGGA 2152
Qy 661 GluThrGlyLysArgSerLysGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 CAGACAGGAAGAGAGTAAGAGGTCATAGTTGGAGATGGAAATGAAATCTTTGTA 2212
Qy 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGTGTGAGACTCCGATGAAGATGATAACAGCTTCCTCAACAAACATCTCCACAA 2272
Qy 701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCCAGTCTCTGAATGTGTCAGTTTGTAGACAACACCTTTGTGTAAGAATTCAC 2332

Qy 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
Db 2333 ACTCAGAATCAGAAATCCAGGATGTGAACTCTGGAGGAGAGAGTGTGTCAAAGAGCTC 2392
Qy 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2393 TCTGTGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2449
RESULT 3
US-09-783-732-3
; Sequence 3, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: (EPLIN)
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3650
; TYPE: DNA
; ORGANISM: Homosapien
US-09-783-732-3
Alignment Scores:
Pred. No.: 0 Length: 3650
Score: 3872.50 Matches: 750
Percent Similarity: 99.08% Conservative: 3
Best Local Similarity: 98.68% Mismatches: 6
Query Match: 98.61% Indels: 1
DB: 11 Gaps: 1
US-09-890-549-4 (1-759) x US-09-783-732-3 (1-3650)
Qy 1 MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 102 ATGGAATCATCTCCATTTAATAGACGGCAATGACCTCACTATCATTTAGGGTAAACAGCC 161
Qy 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 162 AAAGAACTTCTCTCTGTCACAGAACAAAGTCATCGCTATTGTGGAATAATTCTCCAAG 221
Qy 41 TyrGlnLysAlaAlaGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 222 TACCAGAAAGCAGCTCAGAAACAAACATGGAGAACAGAGAGTAACACCGAAATCTC 281
Qy 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 282 TCCAGCAGCTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAAAGTGGGAGAACCCAGGG 341
Qy 81 LeuGlyValaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 342 CTGGAGACAGAGTCTCACACAGACTCTTACGGAAACAGCAGCAGCTGAGATTAGGCACAGA 401
Qy 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 402 GCAGACCATCTCTCTGCTCAAGTGACCAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCA 461
Qy 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 462 GAAGAACAAATCCACCCAGATCTAGACTCAGCTCACCTCTCTGAAGCCCTCGTTCAGGGT 521
Qy 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 522 CGATATCCCCACATCAAGGACCGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 581

QY	161	MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn	180
DB	582	ATGGAATAATGCTCTAGGAGAATCCAGGATGAAGATAGAAAAATCAGAAATCAGTGAAAC	641
QY	181	ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet	200
DB	642	ACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAACAGGCTTAAAGATGATG	701
QY	201	PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer	220
DB	702	TTTGGAAAGGTGAACCAACTCAACTAAGATTCTCCGGGCCCAAGCCGGAAGTGCAGT	761
QY	221	GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln	240
DB	762	GGAGCAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCACGCTCAG	821
QY	241	LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro	260
DB	822	TTGTTCATCTCTACATTTGACTCGAGAGAAAATCAGAGTAGACAGAAATCTGGAACCTTCCA	881
QY	261	ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys	280
DB	882	CCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAAGGACAGTGCCTGCCAA	941
QY	281	GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle	300
DB	942	CAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGTGGCAAAATCAAAATT	1001
QY	301	HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln	320
DB	1002	CATAAATGGAGCAAAAGAGAAATGTGCCCCAGGTCCTGAGGCTGCATCACCCTATCAG	1061
QY	321	GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu	340
DB	1062	GAAGGGGAAAAAGATTCTGCAAAATCAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCAA	1121
QY	341	AspAspSer---ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys	359
DB	1122	GATGACTCCCGAGGTGACTCCAGGTTAAGAGTCAAGGTTCAACAGCCTGTCCATCCCAAG	1181
QY	360	ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMet	379
DB	1182	CCACTAGTCCAGATTCCAGAGCCTCAGTCTTTCTGAAGTTCTCTCCCAAGCAATG	1241
QY	380	LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro	399
DB	1242	AAGAAGTTTCAGGCACCTGTCAAGAGAGACCTGCGTGGAAATGTCAGAGAGACAGTCTATCCA	1301
QY	400	MetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyr	419
DB	1302	ATGAGAGCGTCTCTTGGCCAAACACAGAGGTGTTTCACATCAGCTGCTCTCCGTTCCTAT	1361
QY	420	CysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys	439
DB	1362	TGCACCAACAAACCTCAGTCTAGGAACATATGCACTCTTTACATGGAAGAAATCTATTGTAAG	1421
QY	440	ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArg	459
DB	1422	CCTCAGCTTCAATCAACTCTTTAAATCTAAGGCNACTATGATGAAGGCTTTGGGCACAGA	1481
QY	460	ProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAla	479
DB	1482	CCACACAAGGATCTATGGCAAGCAAAAATGAAACCAAGAGATTTTGGAGACCCAGGCC	1541
QY	480	GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAlaProIleAla	499
DB	1542	CAGCTTGCAAAATGCAAGGGAGACCCCTTCACAGCCAGGGGTAGAAAAATGCCCTATTGCT	1601
QY	500	LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu	519
DB	1602	AAGTGGGTGTCTTGGCTGCAAGTATGGAAGCCAGGCGCTCTCTCAGCAGGAGAAAGAA	1661

RESULT 4

```

US-09-783-732-1
; Sequence 1, Application US/09783732
; Publication NO. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3543
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-783-732-1

```

Alignment Scores:
Pred. No.:

Length: 3543

Score: 3603.50 Matches: 695
 Percent Similarity: 99.04% Conservative: 3
 Best Local Similarity: 98.58% Mismatches: 6
 Query Match: 91.76% Indels: 1
 DB: 11 Gaps: 1

US-09-890-549-4 (1-759) x US-09-783-732-1 (1-3543)

QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
 DB 159 AACACCGAAATCTCTCCAGACACTTTAGAAAGGGACCCCTGACTGTGTAAAGAAGAAG 218

QY 76 TyrGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThr 95
 DB 219 TGGGAGAACCCAGGGCTGGACACAGAGCTCTACACAGACTCTCTACGGAACAGCAGCACT 278

QY 96 GluIleArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly 115
 DB 279 GAGATTAGGCACAGACGACCATCTCTCTGCTGAGTGACACAGCCACGCTCTCTCGA 338

QY 116 AlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGlu 135
 DB 339 GCCAAAGCTGACCAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGAA 398

QY 136 AlaLeuValGlnGlyArgTyrProHisLysLysAspGlyGluAspLeuLysAspHisSer 155
 DB 399 GCCCTCGTTACGGCTCGATATCCCCACATCAAGGACGGTGAGGATCTTTAAAGACCACTCA 458

QY 156 ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
 DB 459 ACAGAAAGTAAATAATGGAATGCTCTAGGAGAAATCCAGCATGAGTAGAATAATCA 518

QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysLysLysLysLysLysLysLysLys 195
 DB 519 GAAATCAGTGAAGAACACACATGCTTCGGCAAAATAGAGAAATATAATGTTCCCGCTGAC 578

QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLysLysLysLys 215
 DB 579 AGGCTTAAGATGATGTTTGAGAAAGGTGAACCACTCAAACTTAAGATTCTCCGGGCCCAA 638

QY 216 SerArgSerAlaSerGlyArgLysLysSerGluAsnSerTyrSerLeuAspAspLeuGlu 235
 DB 639 ACCCGAAGTGAAGTGAAGGAGAGATCTCTGNAACACAGTATTCTTAGATGACCTGGAA 698

QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
 DB 699 ATAGCCCAAGGTGAGTTGTCTTACATTTGACTCGAGAAATATGAGAGTAGACGA 758

QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlaLysTyrGln 275
 DB 759 AATCTGGAACTTCCACGGCTCTCGAAACCTCTATAAAGGATCGAAATGGCCCAAGTACCAG 818

QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
 DB 819 GCAGCTGGCTCCAAACAAGACGAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 878

QY 296 GlyGluIleLysLysLysMetGluGlnLysGluAsnValProProGlyProGluVal 315
 DB 879 GCGGAAATCAAAATTCATAAATGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTC 938

QY 316 CysIleThrHisGlnGluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaValArg 335
 DB 939 TGCAATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAGATAGCCTGGCAGTCCGT 998

QY 336 SerThrProAlaGluAspAspSer---ArgAspSerGlnValLysSerGluValGlnGln 354
 DB 999 TCCACCCCTGCCGAAGATGACTCCCCAGGTGACTCCCAAGGTTAAGAGGTGAGGTTCAACAG 1058

QY 355 ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer 374
 DB 1059 CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCCAGTCTTTCTGAAAGTTCT 1118

QY 375 ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln 394

DB 1119 CCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAG 1178

QY 395 LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCys 414
 DB 1179 AAGACAGTCTATCCAATGGAGCGTCTCTTGCCCAACACAGCAGGCTTTTCACATCAGGTGC 1238

QY 415 PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly 434
 DB 1239 TTCGGTTCCTTCTATTGCAACAACAACACTCAGTCTAGGAACATATGTCATCTTTACATGGA 1298

QY 435 ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu 454
 DB 1299 AGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCTAAGGCAACTATGATGAA 1358

QY 455 GlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluIle 474
 DB 1359 GGCTTTGGGCACAGACCACCAAGGATCTTAGGCAAGCAAAATGAAACGAAGAGATT 1418

QY 475 LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGlu 494
 DB 1419 TTGAGAGACACCGCCAGCTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAA 1478

QY 495 AspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSer 514
 DB 1479 AATGCCCTTATTGCTAAGGTGGGTGCTGCTGCTGCAAGTATGGAAGCCAGGGCTCTCTCT 1538

QY 515 GlnGlnLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProPro 534
 DB 1539 CAGCAGAGAGAGAAACAGCAGCTGAAACCAAGAGCTGAGGATCGCTCGGCCACCC 1598

QY 535 ProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysLysMetSerLysPro 554
 DB 1599 CCACCTGAACTTGGAGGTTTCAGGAAGTCTTGAGAGAGGGATCAAAATGTCAAAGCCC 1658

QY 555 LysTrpProGluAspGluLysSerLysProGluValProGluAspValAspLeuAsp 574
 DB 1659 AAATGGCTCTCTCAGACAGCAATCAGCAAGCCGAAAGTTCCTGAGGATGTCTGATCTAGAT 1718

QY 575 LeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAla 594
 DB 1719 CTGAAGAAGCTAAGACGATCTTCTTCAAGAGAAAGAGCCGCCCTTCACTGTAGCA 1778

QY 595 AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProIleArgLys 614
 DB 1779 GCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCCCCCTATCAGGAA 1838

QY 615 GlyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAlaGluArgLys 634
 DB 1839 GGCTGGAGCATGTTCAGACAGAAATGGAAGATCTGTGGGTGGAAGAGTTTGCAGAAAGCAA 1898

QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
 DB 1899 CAAGTGGAAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAA 1958

QY 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlyGlyHisSerLeuGluMet 674
 DB 1959 ACAAGAAATTTAAAGCAGACAGCAGGAGAGAAAGTAAGAAAGTTCATAGTTTGGAGATG 2018

QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeu 694
 DB 2019 GAGNATGAGAAATTTGTAGAAAATGGTCAGACTCCGATGAAGAGATGAACAGCTTCCTC 2078

QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
 DB 2079 AAACAACAATTTCCACAAGAACCCCAAGTTTTTGAATTTGTCGAGTTTTGTACACACACC 2138

QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
 DB 2139 TTTGTGAAGAATTCACACTACTCAGAATCAGAAATCCCAGGATGTGGAACCTTTGGGAGGA 2198

QY 735 GluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGlu 754

Db 2199 CAAGTGTCTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG 2258
QY 755 AspGluAspGluGlu 759
Db 2259 GATGAGGATGAAGAG 2273
RESULT 5
US-09-822-849A-399
; Sequence 399, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fichtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 399
; LENGTH: 2749
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-399
Alignment Scores:
Pred. No.: 7,78e-223 Length: 2749
Score: 2335.00 Matches: 490
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 64.55% Indels: 1
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x US-09-822-849A-399 (1-2749)
QY 269 AspArgMetAlaLysTyrGlnAlaValSerLysGlnSerSerThrAsnTyrThr 288
Db 1 GATCGAATGGCCAAAGTACCAGGAGCTGTGTCCAAACCAAGAGCTCAACCACTATACA 60
QY 289 AsnGluLeuLysAlaSerGlyGluLulleLysIleHisLysMetGluGlnLysGluAsn 308
Db 61 AATGAGCTGAAGCCAGTGTGTGGCGAAATCAAAATTCATAAAATGGAGCA-AAGGAGAA 119
QY 309 ValProProGlyProGluValCysIleThrHisGlnGlnGlyGluLysIleSerAlaAsn 328
Db 120 GTGCCCCCGAGTCTCTGAGGTCTGATCCACCATCAGGAAGGGGAAAGATTCTGCAAA 179
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 180 GAGATAGCTGGCAGTCTGTTCCACCCCTGCGGAAGATGACTCCCGTACTCCAGGTT 239
QY 349 LysSerGluValGlnGlnProValHisProLysProLysProLysProLysProLysProLys 368
Db 240 AAGAGTGAAGTTCAACAGCCTGTCATCCAGCCACTAAGTCAGATTCAGAGCTCC 299
QY 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 300 AGTCTTTCTGAAGTCTCTCCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAG 359
QY 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGln 408
Db 360 ACCTCGTGGAAATGTCAGAAAGACAGTCTATCAATGAGAGCGTCTCTTGGCCCAACAGAG 419
QY 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428

Db 420 GTGTTTTCATCAGCTGCTTCGTTGCTCCATTGCAACAACAACTCAGTCTAGAAACA 479
QY 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 480 TATGCATCTTTACATGGAAGATCTATTGTAAAGCCTCACTTCAATCAATCTTTAAATCT 539
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468
Db 540 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAAAGGATCTATGGCAGAGCAA 599
QY 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 600 AATGAAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGACCCCT 659
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 660 CACAGCCAGGGGTAGAAGATGCCCTTATTGTAAGTGGGTGCTTGGCTGCAAGTATG 719
QY 509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 720 GAAGCCAAAGGCTCTCTCTCAGCAGGAGAGAAACACAGCCAGCTGAACCAAGAGCTG 779
QY 529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGly 548
Db 780 AGGATCGCTGCGCCACCCCTCACTGAAGTTCAGGAAGTCCCTTGGAGGAGGG 839
QY 549 IleLysMetSerLysProLysTyrProProGluAspGluLulleSerLysProGluValPro 568
Db 840 ATCAAAATGTCAAGCCCAATGGCTCTCAGACGAAATCAGCAAGCCGCAAGATTCCT 899
QY 569 GluAspValAspLeuLysLysLeuArgArgSerSerSerSerLysLysGluArgSer 588
Db 900 GAGGATGTCGATCTAGATCTGAGAAAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 959
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 960 CGCCCATTCATGTAGCAGCTTCATTTCAAAGACCTCTGTCAAGACCCCAAACTGTG 1019
QY 609 SerProProLysArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1020 TCCCCACCTATCAGGAAAGGCTGGAGCATGTCTCAGCAGCAGAGTGAAGAGTCTGTGGGTGGA 1079
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1080 AGAGTTCGAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTG 1139
QY 649 GlyLysThrThrTyrGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1140 GGAAAAACAACCTGGCCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGTAA 1199
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1200 GGTCTAGTTTGGAGATGAGAAATGAGAAATCTGTAGAAAATGGTGCAGACTCCGATGAA 1259
QY 689 AspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTyrSer 708
Db 1260 GATGATAACAGCTTCCTCAAAACAACAAATCTCCACAAGAACCCCAAGTCTCTGAATGGTCG 1319
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrClnAsnGlnLysSerGlnAsp 728
Db 1320 AGTTTTGTAGCAACACCTTTGCTGAAGAAATTCACACTACTCAGAATCAGAAATCCAGGAT 1379
QY 729 ValGluLeuTyrGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1380 GTGGAACCTCGGGAGGAGAGTGTCTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 1439
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1440 AATCGGTATTATGATCAGGATGAGGATGAAGAG 1472
RESULT 6
US-09-925-297-152

```

; Sequence 152, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (729)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-152

```

```

Alignment Scores:
Pred. No.: 6,36e-106 Length: 732
Score: 1254.00 Matches: 238
Percent Similarity: 98.76% Conservative: 0
Best Local Similarity: 98.76% Mismatches: 3
Query Match: 31.93% Indels: 0
DB: 9 Gaps: 0

```

US-09-890-549-4 (1-759) x US-09-925-297-152 (1-732)

```

Qy 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 2 GAAGGGGAAGATTCTGCAAAAGAGATAGCTGGCAGTCCCTTCACCCCTGCCGAA 61
Qy 341 AspAspSerArgAspSerGlnValIysSerGluValGlnGlnProValHisProLysPro 360
Db 62 GATGACTCCCGTACTCCAGGTTAAGAGTGAGGTTCACAGCCGTGTCATCCCAAGCCA 121
Qy 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 122 CTAAAGTCAGATTCCAGAGCTCCAGTCTTTCTGARAGTTCTCTCCCAAGCAATGAAG 181
Qy 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyProMet 400
Db 182 AAGTTTCAGGCACCTGCGAGAGACCTGCTGGAAATGTCAGAAGACAGTCTATCCAAATG 241
Qy 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrcys 420
Db 242 GAGCGTCTCTTGGCCAAACCAGCAGGTGTTTCACATCAGCTGCTTCGTTCTCTATTGC 301
Qy 421 AsnAsnLysLeuSerLeuGlyThrTyAlaSerLeuHisGlyArgIleTyrcysLysPro 440
Db 302 AACCAACAACTCAGCTAGGAACAATATGCATCTTTACATGGAAGAACTATTTAGACCT 361
Qy 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrcysPheGlyHisArgPro 460
Db 362 CACTTCAATCACTCTTTAAATCTAAGGCAACTATGATGAAGGCITTTGGGCACAGACCA 421
Qy 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGln 480
Db 422 CACAAGGATCTATGGCAAGCAAAAATGAAACGAAGAGATTTTGGAGAGACCGCCAG 481
Qy 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 482 CTTGCAATCAAGGGAGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAG 541
Qy 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 542 GTGGGTCTCTKKTGTCAGTATGGAAGCCAGGCCCTCTCTCAGCAGGAGGAAGAGAC 601

```

```

Qy 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGlySer 540
Db 602 AAGCCAGCTGAAACCAAGAGCTGAGATCGCTGCGCCACCCCTGAACCTGGAAGT 661
Qy 541 SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProGluAsp 560
Db 662 TCAGGAAGTGCTTGGAGGAAGGATCAAAATGTCAAGCCCAATGGSTCTCTGAAGAC 721
Qy 561 Glu 561
Db 722 GAA 724

```

RESULT 7

```

US-10-198-846-10895
; Sequence 10895, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10895
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 691, 692, 693, 694, 695, 696, 697, 698
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10895

```

Alignment Scores:

```

Pred. No.: 4,74e-68 Length: 698
Score: 841.50 Matches: 169
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 90.37% Mismatches: 2
Query Match: 21.43% Indels: 16
DB: 15 Gaps: 1

```

US-09-890-549-4 (1-759) x US-10-198-846-10895 (1-698)

```

Qy 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 174 ATGGAATCATCTCCATTAAATAGACGGCAATGGACCTCACTATCATTCATGGGTAAACGCC 233
Qy 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 234 AAAGAACTTCTCTGTCTCAACAAGAACCAAGTCATCGGCTATTGTGAAATATTCTCCAA- 292
Qy 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 293 -----AACACCGAAATCTC 307
Qy 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 308 TCCAGAGCACTTTAGAAAGGGACCCCTGACTGTGTTAAAGAGAGAGTGGGAGAACCCAGGG 367
Qy 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 368 CTGGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGA 427
Qy 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120

```

Db 428 GCAGACCATCTCTCTGCTGAAGTACAGGCACCTCTCTCTGAGCCAAAGCTGACCAA 487
Qy 121 GluGluGlnIleHisProArgSerArgLeuArgSerProGluAlaLeuValGlnGly 140
Db 488 GAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCGTTCAAGGT 547
Qy 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 548 CGATATCCACATCAGGACGGTCTAGGATCTTAAGACCACCTCAACAGAAAGTAAAAAA 607
Qy 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluAsn 180
Db 608 ATGAAATTTGCTAGGAGATCCAGGCATGAGAGTAGAAGTAAAGAAATCAGAAATCAGTGAAC 667
Qy 181 ThrAspAlaSerGlyLysIle 187
Db 668 ACAGATGCTTTCCGGCAATA 688

RESULT 8

US-10-060-036-4338

; Sequence 4338, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4338
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 415..418, 556
; OTHER INFORMATION: n = A,T,C or G

US-10-060-036-4338

Alignment Scores:
Pred. No.: 6,17e-67 Length: 565
Score: 828.00 Matches: 162
Percent Similarity: 98.78% Conservative: 0
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 21.08% Indels: 1
DB: 15 Gaps: 0

US-09-890-549-4 (1-759) x US-10-060-036-4338 (1-565)

Qy 597 PheGlnSerThrSerValLysSerProLysThrValSerProPheIleArgLysGlyTyr 616
Db 1 TTTCAGACCATCTCTGTCAGAGCCCAAAACTGTCTCCACCTATCAGGAAAGGCTGG 60
Qy 617 SerMetSerGluLysSerGluSerValGlyArgValAlaGluArgLysGlnVal 636
Db 61 AGCATGTCTCAGACGAGTGAAGTCTGTGGTGGAGAGTTGCAGAAAGGAACNAGTG 120
Qy 637 GluAsnAlaLysAlaSerLysLysAsnValGlyLysThrThrThrGluAsnLys 656
Db 121 GAAATGCAAGGCTCTTAAGAAAGTGGAAATGTGGAAAAACAACCTGCGCAAAACAA 180
Qy 657 GluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsn 676
Db 181 GAATCTTAAGGAGACAGGAGAGAGAGTAAAGAGTATAGTTTGGAGATCGGAAT 240

Qy 677 GluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGln 696
Db 241 GAGAATCTTTGTAGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACAA 300
Qy 697 GlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAla 716
Db 301 CAATCTCCACAAGAACCCCAAGTCTCTGAATTTGTCGAGTTTGTAGACAACACCTTTGCT 360
Qy 717 GluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly-GluVa 736
Db 361 GAAGAATTCACCTACTACTCAGAAATCCAGGATGTGGAACTCTGGAGGAGGAGNAGT 420
Qy 736 lValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTrpAspGluAspG1 756
Db 421 GGTCAAGAGGCTCTCTGTGGAAGAACAGATAAGAGAAATCGGTATTATGATGAGGATGA 480
Qy 756 uAspGluGlu 759
Db 481 GGATGAAGAG 490

RESULT 9

US-10-094-749-1019
; Sequence 1019, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEXI, NAHOKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1019
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-094-749-1019

Alignment Scores:

Pred. No.: 7.8e-46 Length: 2710
Score: 608.50 Matches: 204
Percent Similarity: 42.42% Conservative: 115
Best Local Similarity: 27.13% Mismatches: 279
Query Match: 15.50% Indels: 154
DB: 13 Gaps: 27

US-09-890-549-4 (1-759) x US-10-094-749-1019 (1-2710)

Qy 12 ThrSerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysSer 31
Db 681 ACTTTTATCAATAGATCTACTGCCCGACGAAACCCCTACTAAGAACGAGCTT 740

QY 32 SerAlaIleValGluIlePheSerLysTyrGlnLysAlaAAlaGluThrAsnMetGlu 51
|||
Db 741 TCTCAGTCCCTAAAGACAGT---TATGTTGAACCCCCACCAGAGGCCCATGTCG 797
QY 52 LysLys-----ArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGly 67
|||
Db 798 CAATAATCTGAATTCACAGACCAACACTTCCCTTCTCCACCAGAGTCTGCTCTGAA 857
QY 68 ThrLeuThrValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla----- 83
|||
Db 858 CAACCTGTGAGACTCAAGACACCACTGCAAAAGTTATCCAAAGGGCCATCCCATGTCCA 917
QY 84 -----GluSerHisThrAspSerLeuArgAsnSerSerThr 95
|||
Db 918 GCAGCAACCCCGTTCCTAATGAGAGAGGCTCTGAAATCATCATGCTCTCCCTGCAACA 977
QY 96 -----GluIleArgHisArgAlaAspHisProProAlaGluValThr--- 109
|||
Db 978 CTTCGTGCTCAATTAAGATAGAACTCGTGTAGGACTCTCCACCTACCAATCACAAATA 1037
QY 110 -----SerHisAlaIleSerGlyAla-----LysAlaAspGlnGlu 122
|||
Db 1038 CCAGTAAATATAAATCATGCTCTAGTGGTTCCTTCAGAGAATCTGTGACGCTCAAGAG 1097
QY 123 GlnIleHisProSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTyr 142
|||
Db 1098 GAAATCAGGAAGTGGAGAAG-----AGAGCTACTTAT 1130
QY 143 ProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetGlu 162
|||
Db 1131 GTTCAT---AAGATGGACTAAATCCACTGATCATGTCGCGGACACTGAAAGTTAT 1187
QY 163 AsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAsp 182
|||
Db 1188 GATGAGTGTAAATCATCCGCAAGTTCGAGTGCCTCTCCCTGTGAGACACACACAG 1247
QY 183 -----AlaSerGlyLysIleGluLysTyrAsn 191
|||
Db 1248 AGATATGAAGGGCCCAACCGAAGTGTCCAAATGGCTGAAATTTCTGTAATGACCCCTGAA 1307
QY 192 ValProLeuAsnArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIle 211
|||
Db 1308 AATGAATTAACAGTGTCTCAGGAATTTGAGCATGGCCCGCATTTCTGGAAGCAAGTCA 1367
QY 212 ---LeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSer 230
|||
Db 1368 AATAGACAGTATTATGCAAGGGAGAAACAAACCATACATACACAGAAAGTCTGTACA 1427
QY 231 LeuAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLys 250
|||
Db 1428 TTTTGTAAAGGAGAAATTTGGATTAAACATCTTTAGGAAACACAGCTTTTACAGACTTTTCT 1487
QY 251 AsnGluSerArgArgAsnLeu-----GluLeuProArgLeu---Ser 263
|||
Db 1488 TGCAACATCTCAGAACTCCGAGAAAGATTCCTGTTAAGACAGCCCGAGATCTGCTCT 1547
QY 264 GluThr---SerIleLysAspArgMetAla-----LysTyrGlnAlaAla 277
|||
Db 1548 GAAACCACTCTCAGAGTGAACATTTCTCAGGCATGGATGCAATTTGAGAGTCAAAATGTT 1607
QY 278 ValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGlu 297
|||
Db 1608 GAGTCGAAGATGAAACCTCTTTCATCATAGCTCAGAGTGCAGAAATCTGCTGTGAC 1667
QY 298 IleLysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIle 317
|||
Db 1668 TTCAGCTT-----GCCCAACCACTATGAGGATGTCATT 1703
QY 318 ThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThr 337
|||
Db 1704 GCT-----GGACATATTTAGATATCTCTGATTTCCCTTAAGAAAGTAAGAAAAAT 1754
QY 338 -----ProAlaGluAsp 341
|||

Db 1755 TTTCAAAGACGTCGGCAAGAGAGTGAAGAGTGTAAAAAGCCTGGGATATGCAACCGCA 1814
QY 342 AspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeu 361
|||
Db 1815 GATGCTTCTCAACTGAGATGAGAACCCCTTCCAAGAG----- 1853
QY 362 SerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLys 381
|||
Db 1854 -----GAATCTGCATTTATAAGTGAAGCTGCT----- 1880
QY 382 PheGlnAlaProAla-ArgGluThrCysValGluCysGlnLysThrValTyrProMetGlu 401
|||
Db 1881 -----GCTCCAAGACAAGGAAATATGTATCTTTGTCAAAAGACAGTATTATCCAATGGA 1934
QY 401 uArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCysAs 421
|||
Db 1935 GTGCCCTAGTGGCAGACAAGCAGAAATTTTCATAAGTCTGCTCCGATGCCACCATTCGAA 1994
QY 421 nAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHi 441
|||
Db 1995 CAGTAACTAAGTTTGGGAAATTTATGCATCATCTTATGGACAAATATATCTGTAAACCTCA 2054
QY 441 sPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHi 461
|||
Db 2055 CTTTAAACAACCTTTTCAAATCCAAAGGAATTTATGATGAAGGTTTTTGGACATAANGCAGCA 2114
QY 461 sLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIle-----LeuGluAr 477
|||
Db 2115 TAAAGATAGATGGAAGTGCACAAACCAAGCAGATCAGTGGACTTTTATCTTCAATCAAGA 2174
QY 477 gProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaPr 497
|||
Db 2175 ACCAAATATGTGTAAATATATTGCAGAAACACCCCTGTACCTGGAGATCGTAATAAGCA 2234
QY 497 oIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnG 517
|||
Db 2235 TTTA-----GATGCTGTGTAACAGTGAAGGGCA 2261
QY 517 uLysGluAsp-----LysProAlaGluThrLysLysLeuArgIleAlaTrpProProPr 535
|||
Db 2262 AAGGAATGATTTGAGAAATTTAGGGGAAAGGGGAAATTTAAAGTCATTTGCCCTCTTC 2321
QY 535 oThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLy 555
|||
Db 2322 CAAGGAGATCCCTTAAGAAACCTTACCTTTTGGAGAGAGCTCAAAATGAGTAAACCTAA 2381
QY 555 sTrpProProGlu---AspGluIleSerLysProGluValProGluAspValAspLeuAs 574
|||
Db 2382 GTGGCCACTGAAATGACAAACCTGCTATCCCTGAAATTTAAAGTGAATCTCTCTCTAGA 2441
QY 574 pLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAl 594
|||
Db 2442 AGATGTTAGAACTCCAGAAATATAAGGACAAAGACAAAGATCACTTTCCATTT---TTGCA 2498
QY 594 alaSerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgly 614
|||
Db 2499 GCCTTATCTACAGTCCACCCATGTTGT----- 2526
QY 614 sGlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAlaGluArgly 634
|||
Db 2527 -----CAGAAAGAGGATGTTATAGGA-----ATCAAGAAAGAAATGAA 2561
QY 634 sGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlu 654
|||
Db 2562 AATGCTTGAAGGAAAGAAAGATGAAAGAAAGAA-----GGAAGGAAAGATGTGCA 2612
QY 654 nAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMe 674
|||
Db 2613 AGATAGCCGAGTGAAGCTGAAGACACAAAGAGTAAACAGGAAAGTGTCTATGGATCTCTAA 2672
QY 674 tGluAsnGluAsnLeuValGluAsnGlyAlaAsp 685
|||

Db 2673 TGACACAATAATGTGTAATGTGACAGTGTCTGAA 2706

RESULT 10

US-09-833-381-1727

; Sequence 1727, Application US/09833381

; Patent No. US20020132090A1

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs

; FILE REFERENCE: 5800-119

; CURRENT APPLICATION NUMBER: US/09/833,381

; CURRENT FILING DATE: 2001-04-11

; PRIOR APPLICATION NUMBER: 09/516,448

; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 2050

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1727

; LENGTH: 3236

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(3236)

; OTHER INFORMATION: n = A,T,C or G

US-09-833-381-1727

Alignment Scores:

Pred. No.:	1..118-45	Length:	3236
Score:	608.00	Matches:	195
Percent Similarity:	42.88%	Conservative:	112
Best Local Similarity:	27.23%	Mismatches:	260
Query Match:	15.48%	Indels:	149
DB:	10	Gaps:	25

US-09-890-549-4 (1-759) x US-09-833-381-1727 (1-3236)

QY 51 GluLysLysArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThr 70

Db 113 GAAATTCACAGAGCAAACTTCCCTTCTCCACCCAGGAGTGTCTGAAACAACCTTGT 172

QY 71 ValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla 83

Db 173 AGACTCAAGACACCACCTGCAAGTATCCAAAGGGGCCATCCCATGTCCAGCAGCAACC 232

QY 84 -----GluSerHisThrAspSerLeuArgAsnSerThr----- 95

Db 233 CCGTTTCCAATTGTAGAGAGAGGCTGAAATCATCATGTCTCTCGCAACACTTCGTCT 292

QY 96 -----GluLeuArgHisArgAlaAspHisProProAlaGluValThr----- 109

Db 293 CAATTAAGATAGAAACTCGTGTGGAGCTCTCCACCTACCAATCACAATACCAAGTAAAT 352

QY 110 ---SerHisAlaAlaSerGlyAla-----LysAlaAspGlnGluGlnHis 125

Db 353 ATAAATCATGTCTAGTGGTTCCTTCAGAGAAATCTGTGGAGCTCAAGAGGAATCAGG 412

QY 126 ProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTyrProHisIle 145

Db 413 AAAGTGGAGAAG-----AAAGTACTTATGTTTCAT--- 442

QY 146 LysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetGluAsn:CysLeu 165

Db 443 AAAGATGCATTAATTCCTCATGATCATGTGGTCCGACACTGAAAGTATGATGCAGTT 502

QY 166 GlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAsp----- 182

Db 503 GAAATCATCCCAAGGTTCAGTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562

QY 183 -----AlaSerGlyLysIleGluLysTyrAsnValProLeu 194

Db 563 GCGGCCAACCGAAGTGTCAATCGCTGCAAAATTCGTGAATGACCTGAAATGAAATA 622

QY 195 AsnArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIle---LeuArg 213

Db 623 AACAGATGGTTCAGGGAATTTGACATGGCCCGAGTTTCTGAAGCAAAAGTCAAAATAGAAGA 682

QY 214 AlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAsp 233

Db 683 GTTTATGCAAGGGAGAGAAACAACATAACAACAAGAAAGTCGTACATTTTGTAAAG 742

QY 234 LeuGluLeuGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSer 253

Db 743 GAGGAATTTGGATTAAACATCTTTAGGAAACACGAGTTTACAGACTTTCTTTCGCAACAT 802

QY 254 ArgArgAsnLeu-----GluLeuProArgLeu---SerGluThr--- 265

Db 803 CCTAGAGAACTGCGAGAAAAGATTCCTGTTTAAGCAGCCCGAGGATCTGCTCTGAACAGG 862

QY 266 SerIleLysAspArgMetAla-----LysTyrGlnAlaAlaValSerLys 280

Db 863 TCTTAAGTGAACATTTCTCAAGCATGGATGCAATTTGAGAGTCAATTTGTTGAGTCCGAG 922

QY 281 GlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300

Db 923 ATGAAAACCTCTTTCATCATCATAGTCTGAGAAAGCTGGCAAAATCTGGCTGTGACTTCAAG--- 979

QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320

Db 980 -----CATGCCCCACCAACCTATGAGGATGTCTATTGCT----- 1012

QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThr----- 337

Db 1013 ---GGACATATTTTATATATCTCTGATTACCTTAAAGAAAGTAAAGAAAAATTTTCAAAG 1069

QY 338 -----ProAlaGluAspAspSerArg 344

Db 1070 ACGTGGCAAGAGAGTGAAGAGATTTTAAAGGCTGGGATATGCAACCCGAGATGTCTTCT 1129

QY 345 AspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeuSerProAsp 364

Db 1130 GCAACTGAGATGAGAAACCTTCCCAAGAG----- 1159

QY 365 SerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAla 384

Db 1160 ---GAATCTGCATTTTATAAGTGAAGCTGCT-----GCT 1189

QY 385 ProAla-ArgGluThrCysValGluCysValGlnLysThrValTyrProMetGluArgLeu 404

Db 1190 CCAAGACAGGAAATATGATATCTTTGTTCAAGAGACAGTTTATCCAAATGGAGTGTCTAG 1249

QY 404 uAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysIle 424

Db 1250 GGCAGACAAGCAGAAATTTTCATAAGTCTCTGCTCCGATGCCACCATGTCACAGTAACT 1309

QY 424 uSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlu 444

Db 1310 AAGTTTGGGAAATATGTCATCACTTCATGGACAAATATATCTGTAAACCTCCTTTAAACA 1369

QY 444 nLeuPheLysSerLysGlyAsnTyrAspGluCysGlyPheGlyHisArgProHisLysAsp 464

Db 1370 ACTTTTCAAAATCCAAAGGAAATTTATGATGAGGTTTGTGACATTAAGCAGCATAAAGATAG 1429

QY 464 uTrpAlaSerLysAsnGluAsnGluIle-----LeuGluArgProAlaGlu 480

Db 1430 ATGGAACTCCAAAACCAAGCAGATCAGTGCAGCTTTTATTCCTTAATGAAGAACCAATAT 1489

QY 480 nLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500

Db 1490 GTGTAATAATATTGAGAAAAACACCTTGTCTCTGGAGATCGTAAATGAACATTTA----- 1544

QY 500 sValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAs 520

Db 1545 -----GATGCTGCTAAACAGTGAAGGCGCAAGGAATGA 1576

QY 520 p-----LysProAlaGluThrLysLysLysLysLysLysLysLysLysLysLysLys 538

Db 1577 TTGAGAAAATTAGGGAAAGGGGAAATTTAAAGTCATTTCCTCCCTCCCAAGGAGAT 1636
Qy 538 uGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProR 558
Db 1637 CCCTAAGAAAACCTTACCCTTTGAGGAGAGAGCTCAAAATGATGATAAAGTGGCCACC 1696
Qy 558 oGlu---AspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLys 577
Db 1697 TGAATGACAAACCTGCTATCCCTGATTTAAAGTGAATCTCTGCTAGAGATGTTAG 1756
Qy 577 sLeuArgSerSerSerLeuLysGluArgSerA-gProPheThrValAlaAlaSerPh 597
Db 1757 AACTCCAGAAATAAAGGACAAAGACAAAGATCACTTTCCATTT---TTGCAGCCTTATCT 1813
Qy 597 eGlnSerThrSerValLysSerProLysThrValSerProLysArgLysGlyTrpse 617
Db 1814 ACAGTCACCCATGTTTGT----- 1832
Qy 617 rMetSerGluGlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValgl 637
Db 1833 -----CAGAAAGAGATGTTATAGGA-----ATCAAGAAATGAAATGCTCGA 1876
Qy 637 uAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysgl 657
Db 1877 AGAAGAAAAGATGAAAAGAAAGAA-----GGAAGGAAGATGTGCAAGATAGGCC 1927
Qy 657 uSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGl 677
Db 1928 GAGTGAAGCTGAAGACACAAAGAGTAAACAGGAAAGTGTCTATGATCTTTAATGACAAACA 1987
Qy 677 uAsnLeuGluAsnGluValAlaAspSerAspGluAspAspAsnSer 692
Db 1988 TAATGTGATGTGTCAGAGTGTGAAAGAGGAGAGAAATGAAAAAACT 2033

RESULT 11

US-09-909-567B-2
; Sequence 2, Application US/0909567B
; Publication No. US20030022257A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A.
; APPLICANT: Nair, Manoj
; APPLICANT: Chen, Seiyu
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes
; FILE REFERENCE: DEX-0214
; CURRENT APPLICATION NUMBER: US/09/909,567B
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/219,834
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 3825
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (430)..(430)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1997)..(1998)
; OTHER INFORMATION: a, c, g, or t
US-09-909-567B-2

Alignment Scores:
Pred. No.: 1.58e-44 Length: 3825
Score: 596.50 Matches: 203
Percent Similarity: 42.50% Conservative: 120
Best Local Similarity: 26.71% Mismatches: 282

Query Match: 15.19% Indels: 155
DB: 11 Gaps: 27
US-09-890-549-4 (1-759) x US-09-909-567B-2 (1-3825)
Qy 12 ThrSerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysAsnLysSer 31
Db 87 ACTTTTATCAAAATAGAAATCTACTGCCCGCAGCAACAGAAAAACCCCTACTTAAGAACGAGCTT 146
Qy 32 SerAlaIleValGluIlePheSerLysLysTyrGlnLysAlaAlaGluGluThrAsnMetGlu 51
Db 147 TCTCAGTCCCTAAAGGACAGT---TATGTTGAACCCCAAGAGGCCCATGTCTG 203
Qy 52 LysLys-----ArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGly 67
Db 204 CAAAAATCTGAAATTCACAGAGCAACACTTCCCTTCTCCACCCAGGAGTCGCTCTGAA 263
Qy 68 ThrLeuThrValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla----- 83
Db 264 CAACCTGTGACAGACTCAAAAGACACCACCTGCAAAAGTTATCCAAAGGGGCCATCCCATGTCCA 323
Qy 84 -----GluSerHisThrAspSerLeuArgAsnSerSerThr 95
Db 324 CGAGCAACCCCGTTCCAAATGTAGAGAGAGGCTTGAATATCATCATGTCTCTCTGCAACA 383
Qy 96 -----GluIleArgHisArgAlaAspHisProPro-AlaGluValThr-- 109
Db 384 CTTGCTGTCGTCAAATTAAGATAGAACTCGTGGTAGGACTCTCCNANCTACATCAATCAAT 443
Qy 110 -----SerHisAlaLysSerGlyAla-----LysAlaAspGlnGluGl 122
Db 444 ACCAGTAATATAAATCATGCTAGTGTGTTCTTCAGAGAAATCTGTGGACGCTCAAGA 503
Qy 122 uGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTy 142
Db 504 GGAATCAGGAAAGTGGAGAG-----AGAGCTACTTA 536
Qy 142 rProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetGl 162
Db 537 TGTTCAT---AAAGATGGACTAAATCCACTGATCATCATGTGTCGCCGACACTGAAAGTTA 593
Qy 162 uAsnCysLeuGlyLysSerArgHisGluValGluLysSerGluIleSerGluAsnThrAs 182
Db 594 TGATGCAGTTGAAATCATCCGCAAGGTTGCAAGTTCCTCGCTGTCAGAGGACACACA 653
Qy 182 p-----AlaSerGlyLysIleGluLysTyAs 191
Db 654 GAGATATGAAGCGGCCCAACCGAACTGTTCAAATGGCTGMAAAATTCGTGAATCACCTGA 713
Qy 191 nValProLeuAsnArgLeuLysMetPheGluLysGlyGluProThrGlnThrLysIl 211
Db 714 AAATGAAATAAACAGATGGTTTCAGGGAATTTGAGCATGGCCCAAGTTTCTGAAGCAAAAGTC 773
Qy 211 e---LeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTySe 230
Db 774 AAATAGAGAGAGTTTATGCAAGGGGAGAAACCAACATAACATAACAACAAGAGTGTGAT 833
Qy 230 rLeuAspAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLy 250
Db 834 ATTTTGAAGGAGGAATTTGGATTAACTCTTTAGGAAACACGAGTTTACAGACTTTTC 893
Qy 250 sAsnGluSerArgArgAsnLeu-----GluLeuProArgLeu---Se 263
Db 894 TTGCAAAACATCTTAGAGAACTCGAGAAAAGATTTCCTGTTAAGCAGCCCGAGGATCTGCTC 953
Qy 263 rGluThr---SerIleLysAspArgMetAla-----LysTyGlnAlaAl 277
Db 954 TGAACACAGGTCTCTAAGTGAACATTTCTCAGGCGATGGATGATTTGAGAGTCAATTTG 1013
Qy 277 aValSerLysGlnSerSerSerThrAsnTyThrAsnGluLeuLysAlaSerGlyGlyGl 297
Db 1014 TGAGTCGAAGATGAAAACCTCTTCATCATCATAGTCTCAGAGCTGCAAAATCTGCTGTGA 1073

QY 297 uilelylleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysI 317
Db 1074 CTTCAAG-----CATGCCCCACCAACCTATGAGGTGTCAT 1109
QY 317 efrHHisGlnGluGlyGlyLysSerAlaAsnGluAsnSerLeuAlaValArgSerTh 337
Db 1110 TGCT-----GGACATATTTAGATATCTCTGATTCACTTAAGAAAGTAAGAAAAA 1160
QY 337 r-----ProAlaGluAs 341
Db 1161 TTTTCAABAGACGTGGCAAGAGAGTGGAGAGTGTTTTAAAGCCTGGGATGATGCAACGC 1220
QY 341 pAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLe 361
Db 1221 AGATGCTTCTGCAACGTAGATGAGAACACCTTCCAAGAG----- 1260
QY 361 userProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMethLysLy 381
Db 1261 -----GAATCTGCATTATTAAGTGAAGCTGCT----- 1287
QY 381 sPheGlnAlaProAla-ArgGluThrCysValGluCysGlnLysThrValTyProMetG 401
Db 1288 -----GCTCCAGACAGGAATATGTATCTTTGTCAAAAGACAGATTATCCAAATGG 1340
QY 401 luArgLeuLeuAlaAsnGlnGlnValPheHisLysSerCysPheArgCysSerTyrcysA 421
Db 1341 AGTCCTAGTGGCAGACAGAGAGATTTTCATAGTCTCTGCTCCGATGCCACCATTTGCA 1400
QY 421 snAsnLysLeuSerLeuGlyThrTyralaSerLeuHisGlyArgIleTyrcysLysProH 441
Db 1401 ACAGTAACTAAGTTTGGGAATATTCATCATCTTCATGACAAATATCTGTAACCTC 1460
QY 441 isPheAsnGlnLeuPheLysSerLysGlyValAsnTyAspGluGlyPheGlyHisArgProH 461
Db 1461 ACTTTAAACAACCTTTTCAATCCAAAGAGAAATATGATGAAGTTTGGACATACAGCAGC 1520
QY 461 isLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIle-----LeuGluA 477
Db 1521 ATAAAGATAGTGAACCTGCAGAAACCAAGACAGATCAGTGGACTTTATTCCTAATGAAG 1580
QY 477 rgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaP 497
Db 1581 AACCAAAATATGTGTAAATAATTTGCAGAAACACCCCTTGACCTGGAGATCGTAATGAAC 1640
QY 497 roIleAlaLysValGlyValLeuAlaSerMetGluAlaLysAlaSerSerGlnGlnG 517
Db 1641 ATTTA-----GATGCTGTGAACAGTGAAGGC 1667
QY 517 lulyLysGluAsp-----LysProAlaGluThrLysLysLeuArgIleAlaTrpProProp 535
Db 1668 AAGGAATGATTTGAGAAATTTAGGGGAAAGGGGAAATTTAAAGTCATTTGGCCTCCTT 1727
QY 535 roThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysLysMetSerLysProL 555
Db 1728 CCAAGGAGATCCCTAAGAAACCTTACCCTTTGAGGAGAGCTCAAAATGATGAACCTTA 1787
QY 555 ystPpProProGlu-----AspGluLysSerLysProGluValProGluAspValAspLeuA 574
Db 1788 AGTGGCCACCTGAAATGACAAACCTCTATCCCTGAATTTAAAGTGAATCTCTGCTAG 1847
QY 574 spLeuLysLysLeuArgArgSerSerLeuLysGluArgSerArgPropPheThrValA 594
Db 1848 AAGATTTAGAACTCCAGAAATAAAGGCAAGACAGATCACTTTCCATTT---TTGC 1904
QY 594 laAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPheArgL 614
Db 1905 AGCCTTATCTACAGTCCACCATGTTGT----- 1933
QY 614 ysglyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAlaGluAgl 634
Db 1934 -----CAGAAAGAGGATGTATAGGA-----ATCAAGAAATGA 1967
QY 634 ysglnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpG 654

Db 1968 AATGCTGAGGAGGAGAAAGATGAAAGNNGAA-----GGAAGGAGAAATGTGTC 2018
QY 654 lnAsnLysGluSerLysGlyGlyThrGlyLysArgSerLysGlyGlyHisSerLeuGluM 674
Db 2019 AAGATAGGCCGAGTGAAGCTGAAGACACAAAGAGTAAACAGGAAAGTCTATGATCTTA 2078
QY 674 etGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSer 692
Db 2079 ATGACACAATAATGATGATTGTCAGAGTGTCTGAAAGGAGAGAAATAAGAAAAAACT 2134
RESULT 12
US-10-161-927-73
; Sequence 73, Application US/10161927
; Publication No. US20030235821A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjal, Tord
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Vernet, Corine
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Gorman, Linda
; APPLICANT: Anderson, David W.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Patturajan, Meera
; APPLICANT: Stone, David J.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: THE SAME
; FILE REFERENCE: 21402-377 D (Cura 677 Other)
; CURRENT APPLICATION NUMBER: US/10161,927
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,575
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/297,414
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 60/297,567
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/298,528
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/325,685
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/299,133
; PRIOR FILING DATE: 2001-06-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 190
; SEQ ID NO 73
; LENGTH: 2955
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1180) .. (2398)
US-10-161-927-73
Alignment Scores: 4.37e-40 Length: 2955
Pred. No.: 4.37e-40 Length: 2955

Score:	546.50	Matches:	182
Percent Similarity:	44.31%	Conservative:	106
Best Local Similarity:	28.00%	Mismatches:	227
Query Match:	13.92%	Indels:	136
DB:	12	Gaps:	24

US-09-890-549-4 (1-759) x US-10-161-927-73 (1-2955)

QY	102	AspHisProProAlaGluValThr-----	-----SerHisAlaAlaSerGlyAla	116
DB	11	GACTCTCCACCTACA--ATCACAATACAGTAATAATAATATCACTCTCTAGTGGTTC	-----	67
QY	117	-----LysAlaAspGlnGluGlnIleHis-ProArgSerArgLeuArgSerPr	133	
DB	68	TTCAGAGAATCTGTGGACGCTCAAGAGGAAATCAGGAAAGTGACGCAAG-	-----	116
QY	133	oProGluAlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAs	153	
DB	117	-----AGAGCTACTATGTTTCAT--AAAGATGGACTAAATTCACCTGA	157	
QY	153	phisSerThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluVal	173	
DB	158	TCACATGTTGCCGACACTGAAGATTATGATCGAGTTGAAATCATCCGCAAGTTGCAGT	217	
QY	173	uLysSerGluIleSerGluAsnThrAsp-----	-----	182
DB	218	GCCTCTCGCCTGTGCAGAGCACACAGAGATATGAAGCGGCCAACCGAACTGTTCAAA	277	
QY	183	-AlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGl	202	
DB	278	GGCTGAAATTTCTGTAATGACCCTGAAATGAAATAAACAGATGGTTTCAGGGAATTTGA	337	
QY	202	uLysGlyGluProThrGlnThrLysIle--LeuArgAlaGlnSerArgSerAlaSerGl	221	
DB	338	GCATGCCCGAGTTTCTGAAGCAAGTCAAAATAGAGAGATTATGCAAGGGGAGAAACAA	397	
QY	221	YArgLysIleSerGluAsnSerTyrSerLeuAspSerLeuGluIleGlyProGlyGlnLe	241	
DB	398	CCATAACATACAACAAAGAAAGTCGTACATTTGT-AAGCAGGAAATTTGGATTAAACATCTT	456	
QY	241	userSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeu-----	257	
DB	457	AGGAACACGAGTTTTCAGACTTTTCTTGCAACATCTCTAGAGAACTCGCAGAAAGAGAT	516	
QY	258	-----GluLeuProArgLeu--SerGluThr--SerIleLysAspArgMetAla--	272	
DB	517	TCCTGTTAAGCAGCCCGAGATCTGCTCTGAAACACAGGTCTCTAAGTGAACATTTCTCAGG	576	
QY	273	-----LysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrTh	288	
DB	577	CATGGATGTCATTGAGAGTCAAAATTTGTTAGTCGAAGATGAAACCTCTTCATCACATAG	636	
QY	288	rAspGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAs	308	
DB	637	CTCAGAAGCTGGCAATCTGCTGTGACTTCAG-----	672	
QY	308	nValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAs	328	
DB	673	TGCCCCACCAACCTATGAGGATGTCATTGCI-----GGACATATTTAGATACTC	723	
QY	328	nGluAsnSerLeuAlaValArgSerThr-----	337	
DB	724	TGATTACCTAAAGAGTAAGAAAAAATTTTCAAAAGACGTGGCAAGAGTGGGAAGAGT	783	
QY	338	-----ProAlaGluAspAspSerArgAspSerGlnValLysSerGluVa	352	
DB	784	TTTTAAAGGCTGGGATATGCAACCCGAGATGCTTCTGCAACTCGAGATGAGAACCACTT	843	
QY	352	lGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGl	372	
DB	844	CCAAGAG-----GAATCTGCATTTTATAGTGA	870	
QY	372	uSerSerProProLysAlaMetLysLysPheGlnAlaProAla-ArgGluThrCysValG	392	

RESULT 13
US-09-880-192-13
; Sequence 13, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.

```

611 AAGCAGATCAGTGGACCTTTATTCCTAATGAAGACCAAAATATGTGTAAAAATATTGCAGA 670
QY uThrProHisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAl 506
Db AAACACCTTCTACTCGAGATCGTAAATGAACATTTA 670
QY aSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 506
Db -----CATGCTGTAAACAGTGAAGGCGCAAGGAATGATTGAGAAAATATTAGGGGA 757
QY uThrLysLysLeuArgIleAlaTTPProProThrGluLeuGlySerSerGlySerAl 544
Db AAGGGGAAAAATTAAGATCATTTGGCCTCTCTCCCAAGGAGATCCCTAAGAAAACCTTACC 817
QY aLeuGluGluGlyLysMetSerLysProLysTrpProGlu 544
Db CTTTGAGGAGAGAGCTCAAAATGAGTAACCTAAGTGGCCACCTGAAATGACAAACCTGCT 877
QY rLysProGluValProGluAspValAspLeuAspLeuLysLysLysLeuArgArgSerSerSe 583
Db ATCCCTCGAATTTAAAGAGTAATCTCTGTAGAAAGATGTTAGAACTCCAGAAAATAAAGG 937
QY rLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerValLy 603
Db ACAAGAGACAGATCACTTCCATTT 937
QY sSerProLysThrValSerProProIleArgLysGlyTrpSerMetSerGluGlnSerGl 623
Db T-----CAGAAAAGA 1003
QY uGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLy 643
Db GGATGTTATAGGA-----ATCAAGAAATGAAAAATGCCTGAAGGAGAAAAGATGAAAA 1057
QY sLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGl 663
Db GAAGGA-----CGAAGGAAGAAATGTCAGATAGCCGAGTGAAGCTGAAGACAC 1108
QY yLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGl 683
Db AAAGAGTAACAGAAAAGTCTATGATCTTAATGACAAACAATATGTGATTGTGCAGAG 1168
QY yAlaAspSerAspGluAspAsnSer 692
Db TGTGAAAAGGAGAAAATGAAAAAACT 1196

RESULT 14
US-10-427-348-13
; Sequence 13, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incvte ID NO. US20030175795A1 050201001

```

```
US-10-427-348-13
Alignment Scores:
Pred. No.: 4,11e-39 Length: 2379
Score: 534.50 Matches: 149
Percent Similarity: 50.22% Conservative: 77
Best Local Similarity: 33.11% Mismatches: 157
Query Match: 13.61% Indels: 69
DB: 13 Gaps: 16

US-09-890-549-4 (1-759) x US-10-427-348-13 (1-2379)
QY 260 ProArgLeu---SerGluThr---SerIleLysAspArgMetAla----- 272
D 1 CCCAGGATGCTCTCTGAACACAGGCTCTTAAGTGAACATTTCTCAGGCATGGATGCAATT 60
QY 273 LysTyGlnAlaAlaValSerLysGlnSerSerThrAsnTyThrAsnGluLeuLys 292
D 61 GAGAGTCAAAATTTGAGTCGAAGATGAAACCTCTCATCATAGCTCAGAGCTGCC 120
QY 293 AlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProGly 312
D 121 AAATCTGGCTGTGATCTCAAG-----CATGCCCCACCAACC 156
QY 313 ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeu 332
D 157 TATGAGATGCTATTGCT-----GGACATATTTAGATATCTCTGATTCACCTAAA 207
QY 333 AlaValArgSerThrProAlaGluAspSerArgAsp-----SerGlnValLysSer 350
D 208 GAAGTAAGAAAAAATTT-TCAAAAGACGTGGCAAGAGAGTGGAGAGTGT-TTAAGGCC 265
QY 351 GluValGlnGlnProValHisProLysProLysProLysProLysProLysSerLeu 370
D 266 TGGGATATGCAACCG-----CAGATGCTTCTGCAACATGAGATGAGAACCACCTTCAA 319
QY 371 SerGluSerSerProLysAlaMetLysLysPheGlnAlaProAla-ArgGluThrCy 390
D 320 GAGGATCTGCA-----TTTATAGTGAAGCTGCTGCCAAGCAAGAAATATG 370
QY 390 sValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnValph 410
D 371 TATACTTTGTCAAAAGACAGTTTATCCATGGAGTGCCTAGTGGCAGACAGCAATTT 430
QY 410 eHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThrVal 430
D 431 TCATAAGTCTCTGCTCCGATGCCACCATTCGAACAGTAACTAAGTTTGGGAAATATGC 490
QY 430 aSerLeuHisGlyArgIleTyCysLysProHisPheAsnGlnLeuPheLysSerLysG 450
D 491 ATCACTTTATGGACAAATATATCTGTAACCTTCAACACTTTTCAAACTTTTCAAAAGG 550
QY 450 yAsnTyArgGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnG 470
D 551 AAATATATGATGAAGGTTTGGACATAACGACGATTAAGATAGATGGAACCTCAAAACCA 610
QY 470 uAsnGluGluLe-----LeuGluArgProAlaGlnLeuAlaAsnAlaArgL 486
D 611 AAGCAGATCACTGGACTTTATTCCTAATGAAGAACCAATATGTGTAATAATATTCAGA 670
QY 486 uThrProHisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAla 506
D 671 AAACACCTTTGATCTGGAGATCGTATGAACATTTA----- 707
QY 506 aSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp-----LysProAlaG 524
D 708 -----CATGCTGTGAACAGTGAAGGCCAAGGAATGATTTGAGAAAAATAGGGGA 757
QY 524 uThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGlySerSerGlySerAl 544
D 758 AAGGGGAAAAATTAAGTCAATTTGGCTCTCTTCCAGGAGATCCTTAAGAAACCTTACC 817
QY 544 aLeuGluGluGlyIleLysMetSerLysProLysProLysProGlu---AspGluIleSe 563
D 818 CTTTGAAGAGAGCTCAAAATGAGTAAACCTTAAGTGGCCACCTGAAATGACCAACCTGCT 877
QY 563 rLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArgSerSerSe 583
D 878 ATCCCTGAATTTAAAGTGAATCTCTCTAGAGATGTTAGAACTCCAGAAAAATAAGG 937
QY 583 rLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerVal 603
D 938 ACAAAGACACAGATCATCTTCCATTT---TTGCACCTTATCTACAGTCCACCCATGTTG 994
QY 603 sSerProLysThrValSerProPheLysArgLysGlyTrpSerMetSerGluGlnSerG 623
D 995 T-----CAGAAAGA 1003
QY 623 uGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLy 643
D 1004 GGATGTTATAGGA-----ATCAAAGAAATGAAATATCCCTGAAAGGAAAGATGAAA 1057
QY 643 sLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrG 663
D 1058 GAAGGAA-----GGAAGGAAGATGTGCAAGATAGCCGAGTGAGCTGAAGACAC 1108
QY 663 yLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnG 683
D 1109 AAAGAGTAACAGGAAAGTGTATGATCTTAATGACAAATAATGTTGTTGTCAGAG 1168
QY 683 yAlaAspSerAspGluAspAspAsnSer 692
D 1169 TGCTGAAAAGGAGAAATAATGAAAAACT 1196

RESULT 15
US-10-440-366-1
; Sequence 1, Application US/10440366
; Publication No. US20030181706A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
; FILE REFERENCE: PB-0006-1 CIP
; CURRENT APPLICATION NUMBER: US/10/440,366
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 09/855,323
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/195,292
; PRIOR FILING DATE: 1998-11-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1298
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 402234CB1
US-10-440-366-1

Alignment Scores:
Pred. No.: 9,75e-26 Length: 1298
Score: 385.00 Matches: 122
Percent Similarity: 42.97% Conservative: 46
Best Local Similarity: 31.20% Mismatches: 137
Query Match: 9.80% Indels: 90
DB: 13 Gaps: 12

US-09-890-549-4 (1-759) x US-10-440-366-1 (1-1298)
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
D 117 CAGAACCCAGCGGGTCCGCTTCTCCACCC-----GAGGCTTCCACCTCCACCGCCA 170
QY 349 LysSerGluValGlnGlnProValHisProLysProLeu---SerProAspSerArgAla 367
D 117 CAGAACCCAGCGGGTCCGCTTCTCCACCC-----GAGGCTTCCACCTCCACCGCCA 170
```

Search completed: January 7, 2004, 01:57:43
Job time : 641 secs

```
Db 171 TGTTCAGGCTGCAGGAGCGCCCGCCAGGCCACCCCTCTCATGACGCCAAAGCGCGGCA 230
Qy 368 SerSerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArg 387
Db 231 GCAGCAGCGTGCAGCGCTCCAGTCTTCAGCGCTG-CGG-----GCCAGGTGAAG 280
Qy 388 GluThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGln 407
Db 281 GAGACCTGCGCGCGCTGCAGAGACCGGTGTACCCCATGGAGCGCTGGTGGCCGACAG 340
Qy 408 GlnValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGly 427
Db 341 CTCATTTTCCCAACTCTGTCTGTCAAGCACTGTACACCAAGCTCAGCGCTGGGC 400
Qy 428 ThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLys 447
Db 401 AGCTACGCGCGCTGCAGCGGAGTCTACTGCAAAACCCCTTCCAGCAGCTGTTTAAG 460
Qy 448 SerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSer 467
Db 461 AGCAAGGCAACTACGACAGGCGTTGGCGCAAGCAGCAGCAGAGGAGCTCTGGGCCAC 520
Qy 468 LysAsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThr 487
Db 521 AAGGAGGTGGAC----- 532
Qy 488 ProHisSerProGlyValGluAspAla----- 537
Db 533 -----CCCGCACCAAGACCGCTGAGCGCTGTAACTTCCACCCCTCTG--- 580
Qy 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
Db 581 -----CGGAGGCTTGGAGCGCGCAGGGGGAAGGT 610
Qy 520 AspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGlu----- 537
Db 611 GGAAGGAGGTGCGAG-----CTGGG-CTTGGTGGGGCCAGG-TGGGAAGGGGAT 659
Qy 538 -----LeuGlySerSerGlySerAlaLeu----- 545
Db 660 GAGCCTTGCTCAGCGGTAGGGGACCGAGCGGCTCTGCTCCAGGACTCCTCTCTT 719
Qy 546 -----GluGluGlyIleLysMetSerLysProLysTyrProProGlu----- 559
Db 720 CTTCTCCCGCAGCGGTGAGGTTGGAAACAGGATTGGGTCTGCCACCCCTGC 779
Qy 560 -----AspGluIleSerLysProGluValProGluAsp-----ValAsp 572
Db 780 TTCTGTCTTCAGCTCCCTCCACCTCACCAGGACCCCTGGAGGCCCCCAA 839
Qy 573 LeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThr 592
Db 840 GCCAGCTCCCTATCTAGGTGCTTTTCTCCAGCAAGGAGTCAGCATGCCCTCAGG 899
Qy 593 ValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPhe 612
Db 900 GTCCAGCTCCCTACTGCCACCGGAGA-CTGTGTGGCCCCCAGCTCTCCCATCTACC 958
Qy 613 ArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGlyArgValAlaGlu 632
Db 959 TCTACCTTAACCTGTTCTGACCCACGAGAGGAGGAGGAGCGCGAGTGGCCAC 1018
Qy 633 ArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsn----- 647
Db 1019 CTGTTGGGATCATATAATGCCCTGAGCCCATGGGGAGGAGATGGGGAAGTGGAGCCA 1078
Qy 648 -----ValGlyLysThrThrTyrGlnAsnLysGlnSerLysGly 660
Db 1079 CCTGCTCTGAGGCAAGGAGGCGCTGCCCTGAGTGGGCTTGGGACCATCTCGAAC 1138
Qy 661 GluThrGlyLysArgSerLysGluGlyHisSer 671
Db 1139 ACCAGGTGGAGAGCAGAGCAAAAGCACTCG 1171
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:51:46 ; Search time 21 Seconds
(without alignments)
1529.235 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPNNRROWTSLSLRVTA.....LSVEQIKRNYDEDEDEE 759

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.1	398	1	US-08-261-822A-16
2	8	1.1	398	5	PCT-US95-07744A-16
3	8	1.1	582	4	US-08-334-179A-4
4	8	1.1	757	4	US-09-252-991A-20231
5	8	1.1	1038	4	US-08-334-179A-2
6	8	1.1	1038	4	US-08-334-179A-8
7	7	0.9	12	3	US-08-844-978-30
8	7	0.9	12	3	US-08-844-978-33
9	7	0.9	15	4	US-08-979-608A-19
10	7	0.9	25	2	US-08-789-078-15
11	7	0.9	25	2	US-08-752-633-16
12	7	0.9	25	3	US-08-844-978-29
13	7	0.9	25	5	PCT-US95-04886-16
14	7	0.9	26	4	US-08-979-608A-20
15	7	0.9	70	4	US-09-252-991A-27697
16	7	0.9	130	4	US-09-252-991A-26359
17	7	0.9	152	2	US-08-608-143-4
18	7	0.9	156	4	US-09-252-991A-28098
19	7	0.9	159	4	US-09-107-532A-3789
20	7	0.9	180	4	US-09-252-991A-21814
21	7	0.9	189	2	US-08-064-694-2
22	7	0.9	189	4	US-09-198-452A-1169
23	7	0.9	200	3	US-08-881-094-19
24	7	0.9	213	4	US-09-252-991A-20163
25	7	0.9	215	4	US-09-252-991A-21524
26	7	0.9	217	4	US-08-979-608A-7
27	7	0.9	240	4	US-09-107-532A-5420

28 7 0.9 251 2 US-08-766-738-1 Sequence 1, Appli
29 7 0.9 251 2 US-08-766-738-3 Sequence 3, Appli
30 7 0.9 251 4 US-09-262-610-1 Sequence 1, Appli
31 7 0.9 251 4 US-09-262-610-3 Sequence 3, Appli
32 7 0.9 290 4 US-09-252-991A-17423 Sequence 17423, A
33 7 0.9 295 4 US-09-252-991A-19816 Sequence 19816, A
34 7 0.9 327 3 US-08-513-9748-372 Sequence 372, App
35 7 0.9 328 3 US-08-459-046-2 Sequence 2, Appli
36 7 0.9 328 3 US-08-513-9748-39 Sequence 39, Appli
37 7 0.9 328 3 US-08-513-9748-371 Sequence 371, App
38 7 0.9 328 4 US-09-102-710B-2 Sequence 2, Appli
39 7 0.9 331 4 US-09-461-436B-39 Sequence 39, Appli
40 7 0.9 338 2 US-09-252-991A-21516 Sequence 21516, A
41 7 0.9 338 2 US-08-933-750C-4 Sequence 4, Appli
42 7 0.9 338 3 US-09-234-613-4 Sequence 322, App
43 7 0.9 365 4 US-09-489-847-322 Sequence 322, App
44 7 0.9 366 4 US-09-252-991A-32385 Sequence 32385, A
45 7 0.9 405 4 US-09-328-352-4349 Sequence 4349, Ap

ALIGNMENTS

RESULT 1
US-08-261-822A-16
; Sequence 16, Application US/08261822A
; Patent No. 5650553
; GENERAL INFORMATION:
; APPLICANT: Ecker, Joseph R. et al.
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; TITLE OF INVENTION: and Pathogens
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5650553ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,822A
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-261-822A-16

Query Match 1.1%; Score 8; DB 1; Length 398;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428

Db 215 NNKLSLGT 222


```
RESULT 2
PCT-US95-07744A-16
; Sequence 16, Application PC/TUS9507744A
; GENERAL INFORMATION:
; APPLICANT: Trustees of The University of Pennsylvania
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07744A
; FILING DATE: 15-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,822
; FILING DATE: June 17, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-07744A-16

Query Match 1.1%; Score 8; DB 5; Length 398;
Best Local Similarity 100.0%; Pred.No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 421 NNKLSLGT 428
Db 215 NNKLSLGT 222

RESULT 3
US-08-334-179A-4
; Sequence 4, Application US/08334179A
; Patent No. 6306622
; GENERAL INFORMATION:
; APPLICANT: ROSENBAUM, JAN S.
; APPLICANT: NOHNO, TSUTOMU
; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
; STREET: 11810 EAST MIAMI RIVER ROAD
; CITY: ROSS
; STATE: OH
; COUNTRY: US
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/334,179A
FILING DATE: 04-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CORSTANJE, BRAHM J.
REGISTRATION NUMBER: 34,804
REFERENCE/DOCKET NUMBER: 5473
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-2858
TELEFAX: 513-627-0260
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 582 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-334-179A-4

Query Match 1.1%; Score 8; DB 4; Length 582;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 329 ENSLAVRS 336
Db 468 ENSLAVRS 475

RESULT 4
US-09-252-991A-20231
; Sequence 20231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20231
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20231

Query Match 1.1%; Score 8; DB 4; Length 757;
Best Local Similarity 100.0%; Pred.No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 LGSERHEV 172
Db 455 LGSERHEV 462

RESULT 5
US-08-334-179A-2
; Sequence 2, Application US/08334179A
; Patent No. 6306622
; GENERAL INFORMATION:
; APPLICANT: ROSENBAUM, JAN S.
; APPLICANT: NOHNO, TSUTOMU
; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
; STREET: 11810 EAST MIAMI RIVER ROAD
```

CITY: ROSS
STATE: OH
COUNTRY: US
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/334,179A
FILING DATE: 04-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CORSTANJE, BRAHM J.
REGISTRATION NUMBER: 34,804
REFERENCE/DOCKET NUMBER: 5473
TELEPHONE: 513-627-2858
TELEFAX: 513-627-0260
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-334-179A-2

Query Match 1.1%; Score 8; DB 4; Length 1038;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 329 ENSLAVERS 336
Db 468 ENSLAVERS 475

RESULT 6
US-08-334-179A-8
Sequence 8, Application US/08334179A
Patent No. 6306622
GENERAL INFORMATION:
APPLICANT: ROSENBAUM, JAN S.
APPLICANT: NOHNO, TSUTOMU
TITLE OF INVENTION: CNA ENCODING A BMP TYPE II RECEPTOR
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
STREET: 11810 EAST MIAMI RIVER ROAD
CITY: ROSS
STATE: OH
COUNTRY: US
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/334,179A
FILING DATE: 04-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CORSTANJE, BRAHM J.
REGISTRATION NUMBER: 34,804
REFERENCE/DOCKET NUMBER: 5473
TELEPHONE: 513-627-2858
TELEFAX: 513-627-0260
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-334-179A-8

Query Match 1.1%; Score 8; DB 4; Length 1038;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 329 ENSLAVERS 336
Db 468 ENSLAVERS 475

RESULT 7
US-08-844-978-30
Sequence 30, Application US/08844978
Patent No. 6075004
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: Peptide Compositions Which Induce Immune
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,978
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 24769-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-844-978-30

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 SYSLDDL 234
Db 4 SYSLDDL 10

RESULT 8
US-08-844-978-33
Sequence 33, Application US/08844978
Patent No. 6075004
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia

APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: Peptide Compositions Which Induce Immune
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,978
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 24769-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: circular
MOLECULE TYPE: peptide
US-08-844-978-33

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 4 SYSDDL 10

RESULT 9
US-08-979-608A-19
Sequence 19, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Atjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,608A
FILING DATE: 28-NOV-1997
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-979-608A-19

Query Match 0.9%; Score 7; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 DEDEDE 759
Db 7 DEDEDE 13

RESULT 10
US-08-789-078-16
Sequence 16, Application US/08789078
Patent No. 5843885
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,078
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,513
FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
US-08-789-078-16

Query Match 0.9%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 11
US-08-752-633-16
Sequence 16, Application US/08752633
Patent No. 5863889
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbets, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
US-08-752-633-16

Query Match 0.9%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 12
US-08-844-978-29
Sequence 29, Application US/08844978
Patent No. 6075004
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbets, Scott
TITLE OF INVENTION: Peptide Compositions Which Induce Immune
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,978
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 24769-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-844-978-29

Query Match 0.9%; Score 7; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 13
PCT-US95-04886-16
Sequence 16, Application PC/TUS9504886
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbets, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
PCT-US95-04886-16

Query Match 0.9%; Score 7; DB 5; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db |||||
3 SYSDDL 9

RESULT 14

US-08-979-608A-20
Sequence 20, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:

APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,608A

FILING DATE: 26-Nov. 6355451-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/048,547

FILING DATE: 03-JUN-1997

APPLICATION NUMBER: US 60/031,930

FILING DATE: 27-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Myers, Louis

REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-08-979-608A-20

Query Match 0.9%; Score 7; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 DEDEEE 759
Db |||||
7 DEDEEE 13

RESULT 15

US-09-252-991A-27697
Sequence 27697, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27697

LENGTH: 70

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27697

Query Match 0.9%; Score 7; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 RSASGRK 223
Db |||||
19 RSASGRK 25

Search completed: January 6, 2004, 09:55:05
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:50:50 ; Search time 38 Seconds
(without alignments)
4007.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPFNRQWTSLSLRTA.....LSVEEIKRNYDEDEDE 759

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 73937 seqs, 200641211 residues

Word size : 0

Total number of hits satisfying chosen parameters: 73937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:

1: /cgn2_5/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	657	86.6	755	11	US-09-783-732-4
2	313	41.2	596	11	Sequence 4, Appli
3	184	24.3	243	9	US-09-783-732-2
4	41	5.4	52	11	US-09-925-297-611
5	15	2.0	527	9	Sequence 611, Appl
6	15	2.0	527	12	US-09-783-732-5
7	14	1.8	127	12	Sequence 53, Appl
8	14	1.8	127	12	US-10-427-348-53
9	14	1.8	158	9	US-10-440-366-12
10	10	1.3	406	12	US-10-104-047-2959
11	8	1.1	236	10	US-09-789-919-42
12	8	1.1	297	12	US-10-161-927-74
13	8	1.1	398	11	US-09-784-864-1047
14	8	1.1	514	12	Sequence 74, Appl
15	8	1.1	1038	10	US-10-094-749-1748
					Sequence 1748, Ap
					Sequence 14037, A
					Sequence 2, Appli

16 1.1 1038 12 US-10-286-152A-42 Sequence 42, Appli
17 8 1.1 1080 9 US-09-904-380-2 Sequence 2, Appli
18 7 0.9 10 11 US-09-572-404B-4124 Sequence 4124, Ap
19 7 0.9 15 9 US-09-962-055-19 Sequence 19, Appl
20 7 0.9 15 10 US-09-976-740-19 Sequence 19, Appl
21 7 0.9 15 14 US-10-023-529-19 Sequence 19, Appl
22 7 0.9 15 14 US-10-023-523-19 Sequence 19, Appl
23 7 0.9 16 15 US-10-225-567A-1216 Sequence 1216, Ap
24 7 0.9 26 9 US-09-962-055-20 Sequence 20, Appl
25 7 0.9 26 10 US-09-976-740-20 Sequence 20, Appl
26 7 0.9 26 14 US-10-023-529-20 Sequence 20, Appl
27 7 0.9 26 14 US-10-023-523-20 Sequence 20, Appl
28 7 0.9 40 11 US-09-764-891-4557 Sequence 4557, Ap
29 7 0.9 53 11 US-09-764-891-4689 Sequence 4689, Ap
30 7 0.9 68 9 US-09-864-761-39951 Sequence 39951, A
31 7 0.9 68 12 US-10-029-386-28019 Sequence 28019, A
32 7 0.9 76 11 US-09-374-046A-100 Sequence 100, App
33 7 0.9 98 15 US-10-106-698-5523 Sequence 5523, Ap
34 7 0.9 154 12 US-10-369-433-11211 Sequence 11211, A
35 7 0.9 161 12 US-10-238-075-12026 Sequence 12026, A
36 7 0.9 181 15 US-10-156-761-10052 Sequence 33, Appl
37 7 0.9 186 14 US-09-934-455-342 Sequence 342, App
38 7 0.9 196 11 US-10-218-137-3 Sequence 3, Appli
39 7 0.9 205 12 US-10-233-448-3 Sequence 7, Appli
40 7 0.9 205 15 US-09-962-055-7 Sequence 7, Appli
41 7 0.9 217 9 US-09-976-740-7 Sequence 7, Appli
42 7 0.9 217 10 US-10-023-529-7 Sequence 7, Appli
43 7 0.9 217 14 US-10-023-523-7 Sequence 7, Appli
44 7 0.9 217 14 US-10-023-523-7 Sequence 7, Appli
45 7 0.9 232 12 US-10-369-433-13799 Sequence 13799, A

ALIGNMENTS

RESULT 1

US-09-783-732-4
; Sequence 4, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783.732
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-4

Query Match 86.6%; Score 657; DB 11; Length 755;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVIFSKYQKAAETNMEKRSNTENL	60
Db	1	MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVIFSKYQKAAETNMEKRSNTENL	60
QY	61	SOHFRKGTITLVKKKWNPCLGAEHTDSLRSNSTEIRADHPPPAEVTSHAASGAKAQD	120
Db	61	SOHFRKGTITLVKKKWNPCLGAEHTDSLRSNSTEIRADHPPPAEVTSHAASGAKAQD	120
QY	121	EQIHPRSLRSPPEALVQGRYPHIKQGEDLKHSTESKMKENCLGESRHEVEKSEISEN	180
Db	121	EQIHPRSLRSPPEALVQGRYPHIKQGEDLKHSTESKMKENCLGESRHEVEKSEISEN	180

```
QY 181 TDASGKIEKYNVPLNRLKMFEXGEBTQTKILRAQSRASGRKISSENSYSLDLEIGPQ 240
Db 181 TDASGKIEKYNVPLNRLKMFEXGEBTQTKILRAQSRASGRKISSENSYSLDLEIGPQ 240
QY 241 LSSSTFDSKSEKSRNLELPRLETSIKORMAKYQAAVSKQSSNTYTNELKASGEIKI 300
Db 241 LSSSTFDSKSEKSRNLELPRLETSIKORMAKYQAAVSKQSSNTYTNELKASGEIKI 300
QY 301 HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSEVQPVHPKP 360
Db 301 HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSEVQPVHPKP 360
QY 361 LSPDSRASSLSSESPPKAMKXFOAPARETCVCQKTVYPMERLLANQQVFHISCFRCSYC 420
Db 361 LSPDSRASSLSSESPPKAMKXFOAPARETCVCQKTVYPMERLLANQQVFHISCFRCSYC 420
QY 421 NNKLSLGTYSASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDLWASKNENEIILERAQ 480
Db 421 NNKLSLGTYSASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDLWASKNENEIILERAQ 480
QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTELGS 540
Db 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTELGS 540
QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTVAASFST 600
Db 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTVAASFST 600
QY 601 SVKSPKTVSPPIRKGMSESESVGGRVAERKOVENAKSKNGNVGKTTWQNK 657
Db 601 SVKSPKTVSPPIRKGMSESESVGGRVAERKOVENAKSKNGNVGKTTWQNK 657
```

RESULT 2

```
US-09-783-732-2
; Sequence 2, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: (EPLIN)
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-2
```

```
Query Match 41.2%; Score 313; DB 11; Length 596;
Best Local Similarity 100.0%; Pred. No. 9.8e-290;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 345 DSQVSEVQPVHPKPLSPDSRASSLSSESPPKAMKXFOAPARETCVCQKTVYPMERLL 404
Db 186 DSQVSEVQPVHPKPLSPDSRASSLSSESPPKAMKXFOAPARETCVCQKTVYPMERLL 245
QY 405 ANQQVFHISCFRCSYCNKLSLGTYSASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDL 464
Db 246 ANQQVFHISCFRCSYCNKLSLGTYSASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDL 305
QY 465 WASKNENEIILERAQANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAE 524
Db 306 WASKNENEIILERAQANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAE 365
```

```
QY 525 TKLURIWPPPTTELGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSL 584
Db 366 TKLURIWPPPTTELGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSL 425
QY 585 KERSRPTVAASFSTSVKSPKTVSPPIRKGMSESESVGGRVAERKOVENAKSKK 644
Db 426 KERSRPTVAASFSTSVKSPKTVSPPIRKGMSESESVGGRVAERKOVENAKSKK 485
QY 645 NGNVGKTTWQNK 657
Db 486 NGNVGKTTWQNK 498
```

RESULT 3

```
US-09-925-297-611
; Sequence 611, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (237)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (238)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (243)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-611
```

```
Query Match 24.2%; Score 184; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 5.7e-167;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 321 EGEKISANENSLAVRSTPAEDDSRDSQVSEVQPVHPKPSPDSRASSLSSESPPKAMK 380
Db 1 EGEKISANENSLAVRSTPAEDDSRDSQVSEVQPVHPKPSPDSRASSLSSESPPKAMK 60
QY 381 KFOAPARETCVCQKTVYPMERLLANQQVFHISCFRCSYCNKLSLGTYSASLHGRYCKP 440
Db 61 KFOAPARETCVCQKTVYPMERLLANQQVFHISCFRCSYCNKLSLGTYSASLHGRYCKP 120
QY 441 HFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIILERAQANARETPHSPGVEDAPIAK 500
Db 121 HFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIILERAQANARETPHSPGVEDAPIAK 180
QY 501 VGVVL 504
Db 181 VGVVL 184
```

RESULT 4

```
US-09-783-732-5
; Sequence 5, Application US/09783732
; Publication No. US20030054417A1
```


GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: 10609/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/659,400
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (0)...(0)
; OTHER INFORMATION: mutant sterol regulatory element binding protein 2
; US-09-783-732-5

Query Match 5.4%; Score 41; DB 11; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.3e-31;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 CQKTVPMERLLANQQVVFHISCFRCSCYCNKLSLGTYSLSH 433
DB 4 CQKTVPMERLLANQQVVFHISCFRCSCYCNKLSLGTYSLSH 44

RESULT 5
US-09-880-192-53
; Sequence 53, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 058201CD1
US-09-880-192-53

Query Match 2.0%; Score 15; DB 9; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
DB 45 QLFKSKGNYDEGFGH 59

RESULT 6
US-10-427-348-53
; Sequence 53, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.

; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 058201CD1
US-10-427-348-53

Query Match 2.0%; Score 15; DB 12; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
DB 45 QLFKSKGNYDEGFGH 59

RESULT 7
US-10-440-366-12
; Sequence 12, Application US/10440366
; Publication No. US20030181706A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
; FILE REFERENCE: PB-0006-1 CIP
; CURRENT APPLICATION NUMBER: US/10/440,366
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 09/855,323
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/195,292
; PRIOR FILING DATE: 1998-11-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 402234CD1
US-10-440-366-12

Query Match 1.8%; Score 14; DB 12; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 457
DB 94 QLFKSKGNYDEGFGH 107

RESULT 8
US-10-104-047-2959
; Sequence 2959, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2959
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2959

Query Match 1.8%; Score 14; DB 12; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNVDEGFG 457
Db 94 QLFKSKGNVDEGFG 107

RESULT 9
US-09-789-919-42
; Sequence 42, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPIETIC BLOOD FORMING STEM
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-42

Query Match 1.8%; Score 14; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNVDEGFG 457
Db 95 QLFKSKGNVDEGFG 108

RESULT 10
US-10-161-927-74
; Sequence 74, Application US/10161927
; Publication No. US2003023821A1
; GENERAL INFORMATION:
; APPLICANT: Zernhosen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjalt, Tord
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Vernet, Corine
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Gorman, Linda
; APPLICANT: Anderson, David W.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Patturajan, Neera
; APPLICANT: Stone, David J.

; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-377 D (Cura 677 Other)
; CURRENT APPLICATION NUMBER: US/10/161,927
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,575
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/297,414
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 60/297,567
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/298,528
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/325,685
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/299,133
; PRIOR FILING DATE: 2001-06-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 190
; SEQ ID NO 74
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-927-74

Query Match 1.3%; Score 10; DB 12; Length 406;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 KMSKPKWPPE 559
Db 62 KMSKPKWPPE 71

RESULT 11
US-09-764-864-1047
; Sequence 1047, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1047
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (44)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (226)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

; LOCATION: (228)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1047

Query Match 1.1%; Score 8; DB 10; Length 236;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 YDEDEDEE 759
|||||
DB 107 YDEDEDEE 114

RESULT 12

US-10-094-749-1748
; Sequence 1748, Application US/10094749
; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOUYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1748

; LENGTH: 297

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-094-749-1748

Query Match 1.1%; Score 8; DB 12; Length 297;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 YDEDEDEE 759
|||||
DB 157 YDEDEDEE 164

RESULT 13

US-09-932-367A-6

; Sequence 6, Application US/09932367A

; Publication No. US20030027152A1

; GENERAL INFORMATION:

; APPLICANT: RHODES, Simon J.

; APPLICANT: BRIDWELL, Jeanne L.

; APPLICANT: MEIER, Bradley C.

; APPLICANT: PARKER, Gretchen E.

; APPLICANT: PRICE, Jeffrey R.

; APPLICANT: SHOWALTER, Aaron D.

; APPLICANT: SLOOP, Kyle W.

; TITLE OF INVENTION: GENERATION OF DIAGNOSTIC TOOLS TO ASSAY THE HUMAN
; TITLE OF INVENTION: LHX3P-LIM/LIM-3 FACTOR
; FILE REFERENCE: 053884-5003
; CURRENT APPLICATION NUMBER: US/09/932,367A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: PCT/US00/04424
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,110
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Danio rerio
US-09-932-367A-6

Query Match 1.1%; Score 8; DB 11; Length 398;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 368 SSLSESSP 375
|||||
DB 272 SSLSESSP 279

RESULT 14

US-10-369-493-14097

; Sequence 14097, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 14097

; LENGTH: 514

; TYPE: PRT

; ORGANISM: Pseudomonas fluorescens

US-10-369-493-14097

Query Match 1.1%; Score 8; DB 12; Length 514;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 SVKSPKTV 608
|||||
DB 487 SVKSPKTV 494

RESULT 15

US-09-908-500A-2

; Sequence 2, Application US/09908500A

; Patent No. US20020102576A1

; GENERAL INFORMATION:

; APPLICANT: James Loyd

; APPLICANT: Kirk B. Lane

; APPLICANT: John A. Phillips, III

; TITLE OF INVENTION: METHOD OF DIAGNOSING PULMONARY

; FILE REFERENCE: 22000.0108U3

; CURRENT APPLICATION NUMBER: US/09/908,500A

; CURRENT FILING DATE: 2001-07-17

; PRIOR APPLICATION NUMBER: 60/216,740

```
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/220,133
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1038
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-908-500A-2

Query Match      1.1% Score 8; DB 10; Length 1038;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      329 ENSLAVRS 336
      |||||
Db      468 ENSLAVRS 475

Search completed: January 6, 2004, 09:54:30
Job time : 38 secs
```

us-09-890-549-4

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:48:00 ; Search time 48 Seconds (without alignments)

2509.864 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSFPNRRQWTSLSRVTA.....LSVREQIKRNYDEDEDEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A. Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*

2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*

3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*

4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*

5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*

6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*

7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*

8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*

9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*

10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*

11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*

12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*

13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*

14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*

15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*

16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*

17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*

18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*

19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*

20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*

21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*

22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*

23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

10 240 31.6 301 23 ABB56420 Human cancer suppr

11 238 31.4 457 23 AAU10979 Human steroid regul

12 227 29.9 232 21 AAB51883 Human secreted pro

13 184 24.2 243 21 AAB54159 Human pancreatic c

14 170 22.4 235 20 AAY73916 Human prostate tum

15 155 20.4 337 22 ABG24090 Novel human diagno

16 131 17.3 476 22 ABG08559 Novel human diagno

17 87 11.5 152 22 AAM40799 Human polypeptide

18 67 8.8 88 21 AAG02073 Human secreted pro

19 44 5.8 331 22 ABG08554 Novel human diagno

20 28 3.7 234 21 AAB51882 Gene 5 human secre

21 28 3.7 299 21 AAB34205 Gene 41 human secr

22 15 2.0 527 23 ABF43758 Human breast tumou

23 15 2.0 527 23 AAE25978 Human protein #5 c

24 15 2.0 547 22 AAU20575 Human secreted pro

25 15 2.0 547 22 AAU20632 Human secreted pro

26 15 2.0 547 23 AAE16626 Human 41441 protei

27 14 1.8 111 22 ABG12835 Novel human diagno

28 14 1.8 127 21 AAY94447 Novel inflammation

29 14 1.8 135 22 ABG12836 Novel human diagno

30 14 1.8 158 21 AAY79176 Haematopoietic ste

31 14 1.8 173 23 ABG06053 Human NS protein s

32 14 1.8 173 23 ABG06054 Human NS protein s

33 14 1.8 173 23 ABG06055 Human NS protein s

34 14 1.8 173 23 ABG06056 Human NS protein s

35 14 1.8 173 23 ABG06057 Human NS protein s

36 14 1.8 173 23 ABG06058 Human NS protein s

37 14 1.8 217 22 AAU20850 Human secreted pro

38 13 1.7 51 21 AAG03479 Human secreted pro

39 11 1.4 90 21 AAB32625 Eucalyptus grandis

40 11 1.4 104 21 AAB32623 Eucalyptus grandis

41 11 1.4 125 23 ABP41582 Human ovarian anti

42 10 1.3 171 22 ABG04347 Novel human diagno

43 9 1.2 9 24 ABG20052 MHC binding peptid

44 9 1.2 283 23 AAO15050 Human 187-2 protei

45 8 1.1 115 23 ABP31959 Human ORF932 prote

ALIGNMENTS

RESULT 1

AA97286

AA97286 standard; Protein; 759 AA.

AC AA97286;

XX

XX 03-JAN-2001 (first entry)

DT

XX Lipid associated protein (LIPAP) 2766980CD1.

DE

XX Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;

XX antagonist; antibody; cardiovascular disease; neurological disease;

KW gastrointestinal disease; lipid metabolism; detection;

KW amplification; monitoring; hybridisation; antisense; triplex;

KW ribzyme; screening; immunoassay.

XX

OS Homo sapiens.

XX

PH Key

FT Location/Qualifiers

FT Modified-site 15

FT /note= "Potential phosphorylation site"

FT Modified-site 19

FT /note= "Potential phosphorylation site"

FT Modified-site 29

FT /note= "Potential glycosylation site"

FT Modified-site 48

FT /note= "Potential phosphorylation site"

FT Modified-site 55

FT /note= "Potential phosphorylation site"

FT Modified-site 59

FT /note= "Potential phosphorylation site"

FT Modified-site 68

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	759	100.0	759	21	AA97286
2	759	100.0	759	22	ABU52869
3	759	100.0	759	22	AAM39013
4	678	89.3	759	22	AAB67701
5	538	70.9	760	21	AAB42934
6	457	60.2	457	22	ABG95477
7	370	48.7	471	20	AA948487
8	334	44.0	600	22	AAB67700
9	299	39.4	299	21	AAB34206

FT	Modified-site	/note= "Potential phosphorylation site"	89
FT	Modified-site	/note= "Potential phosphorylation site"	85
FT	Modified-site	/note= "Potential phosphorylation site"	92
FT	Modified-site	/note= "Potential glycosylation site"	93
FT	Modified-site	/note= "Potential phosphorylation site"	132
FT	Modified-site	/note= "Potential phosphorylation site"	158
FT	Modified-site	/note= "Potential phosphorylation site"	168
FT	Modified-site	/note= "Potential phosphorylation site"	184
FT	Modified-site	/note= "Potential phosphorylation site"	220
FT	Modified-site	/note= "Potential phosphorylation site"	225
FT	Modified-site	/note= "Potential phosphorylation site"	229
FT	Modified-site	/note= "Potential phosphorylation site"	230
FT	Modified-site	/note= "Potential phosphorylation site"	244
FT	Modified-site	/note= "Potential phosphorylation site"	248
FT	Modified-site	/note= "Potential phosphorylation site"	251
FT	Modified-site	/note= "Potential glycosylation site"	253
FT	Modified-site	/note= "Potential phosphorylation site"	266
FT	Modified-site	/note= "Potential phosphorylation site"	286
FT	Modified-site	/note= "Potential glycosylation site"	294
FT	Modified-site	/note= "Potential phosphorylation site"	318
FT	Modified-site	/note= "Potential phosphorylation site"	326
FT	Modified-site	/note= "Potential phosphorylation site"	337
FT	Domain	/note= "Potential phosphorylation site"	344..444
FT	Modified-site	/label= LIM domain	369
FT	Modified-site	/note= "Potential phosphorylation site"	389
FT	Modified-site	/note= "Potential phosphorylation site"	467
FT	Modified-site	/note= "Potential phosphorylation site"	514
FT	Modified-site	/note= "Potential phosphorylation site"	525
FT	Modified-site	/note= "Potential phosphorylation site"	543
FT	Modified-site	/note= "Potential phosphorylation site"	563
FT	Modified-site	/note= "Potential phosphorylation site"	582
FT	Modified-site	/note= "Potential phosphorylation site"	583
FT	Modified-site	/note= "Potential phosphorylation site"	601
FT	Modified-site	/note= "Potential phosphorylation site"	604
FT	Modified-site	/note= "Potential phosphorylation site"	617
FT	Modified-site	/note= "Potential phosphorylation site"	642
FT	Modified-site	/note= "Potential phosphorylation site"	

FT	Modified-site	658		"Potential phosphorylation site"
FT	Modified-site	/note=	662	"Potential phosphorylation site"
FT	Modified-site	/note=	686	"Potential phosphorylation site"
FT	Modified-site	/note=	698	"Potential phosphorylation site"
FT	Modified-site	/note=	706	"Potential phosphorylation site"
FT	Modified-site	/note=	709	"Potential glycosylation site"
FT	Modified-site	/note=	714	"Potential phosphorylation site"
FT	Modified-site	/note=	741	"Potential phosphorylation site"
FT	Modified-site	/note=		"Potential phosphorylation site"
XX				
PN	WO2000049043-A2.			
PD	XZ			
PD	24-AUG-2000.			
PF	18-FEB-2000; 2000WO-USO4160.			
XX				
PR	19-FEB-1999; 99US-0120703.			
PR	08-JUL-1999; 99US-0142762.			
XX				
PA	(INCY-) INCYTE PHARM INC.			
XX				
PI	Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;			
XX				
DR	WPI; 2000-549264/50.			
NR	NP-PSTDB; AAA53826.			
XX				
PT	New human lipid-associated proteins, nucleic acids, and antibodies,			
PT	useful for diagnosis, treatment and prevention of e.g. cardiovascular			
PT	disease			
PS				
PS	Claim 1; Page 76-77; 93pp; English.			
XX				
CC	Lipid-associated proteins (LIPAP) can be used for treating or			
CC	preventing disorders associated with decreased expression of LIPAP,			
CC	for screening for agonists or antagonists of Lirapap, and to raise			
CC	specific antibodies. Antagonists and antagonists of Lirapap are useful			
CC	for treating diseases associated with reduced or increased levels of			
CC	Lirapap, e.g. cardiovascular, neurological and gastrointestinal			
CC	diseases and disorders of lipid metabolism. Fragments of the nucleic			
CC	acid encoding Lirapap are useful for detection of full length coding			
CC	sequences, in hybridization and/or amplification assays or for			
CC	diagnosis or monitoring. Nucleotides encoding Lirapap are used			
CC	to screen for compounds that specifically modify Lirapap expression,			
CC	for recombinant production of Lirapap, in gene therapy, as a source of			
CC	therapeutic antisense, triplex-forming, or ribozyme agents and for			
CC	genomic mapping. Antibodies to the proteins are used for diagnosis			
CC	and monitoring of Lirapap-associated disease by immunoassay, as			
CC	antagonists, in competitive drug screens and for affinity			
CC	purification of natural Lirapap.			

```

AA
SQ      Sequence      759 AA;

Query Match      100.0%;      Score 759;      DB 21;      Length 759;
Best Local Similarity 100.0%;      Pred. No. 0;
Matches 759;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0

QY      1      MESSPFNRORWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL 60
Db      1      MESSPFNRORWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL 60

QY      61      SQHFRKGTLLVLKKKWNPGLGAEHSHTDSLRNSSTETIRHRADHPPEVTSHAASGAKADQ 120
Db      61      SQHFRKGTLLVLKKKWNPGLGAEHSHTDSLRNSSTETIRHRADHPPEVTSHAASGAKADQ 120

QY      121      EEOIHRPSRLRSPPEALVQGRYPHIKDGEDLKHSTSKQKNCNLCGSRHVEKSEISEN 180

```

```

AA
SQ      Sequence      759 AA;

Query Match      100.0%;      Score 759;      DB 21;      Length 759;
Best Local Similarity 100.0%;      Pred. No. 0;
Matches 759;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0

QY      1      MESSPFNRORWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL 60
Db      1      MESSPFNRORWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL 60

QY      61      SQHFRKGTLLVLKKKWNPGLGAEHSHTDSLRNSSTETIRHRADHPPEVTSHAASGAKADQ 120
Db      61      SQHFRKGTLLVLKKKWNPGLGAEHSHTDSLRNSSTETIRHRADHPPEVTSHAASGAKADQ 120

QY      121      EEOIHRPSRLRSPPEALVQGRYPHIKDGEDLKHSTSKQKNCNLCGSRHVEKSEISEN 180

```

Db 121 EQIHPRLRLSPPEALVQGRYPHIKDGEDLKDHSTESKMKENCLGESRHEVEKSEISEN 180
QY 181 TDASKTEKNVPLNRLKMFKEGPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASKTEKNVPLNRLKMFKEGPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSEKNESSRNLELRLSETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGGEIKI 300
Db 241 LSSSTFDSEKNESSRNLELRLSETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGGEIKI 300
QY 301 HKMQEKNVPPGPEVCITHEGEKISANENSLAVRSTPAEDDSDSDSQVKSEVQPVHPKP 360
Db 301 HKMQEKNVPPGPEVCITHEGEKISANENSLAVRSTPAEDDSDSDSQVKSEVQPVHPKP 360
QY 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVCEQKTYVPMERLLANQOVPHISCFCRSCYC 420
Db 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVCEQKTYVPMERLLANQOVPHISCFCRSCYC 420
QY 421 NNKLSLGTYSALHGRYICKPHFNOLFYSKGNVDEGFGHRPHKDLWASKNENEILERPAP 480
Db 421 NNKLSLGTYSALHGRYICKPHFNOLFYSKGNVDEGFGHRPHKDLWASKNENEILERPAP 480
QY 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETTKLRIAMPPTTELGS 540
Db 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETTKLRIAMPPTTELGS 540
QY 541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKRSRPTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKRSRPTVAASFQST 600
QY 601 SVKSPKTVSPPIRGWSMSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRGWSMSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
QY 721 TONOKSQDVELWEGEVVKELSVBEOIKRNYRYDEDEDE 759
Db 721 TONOKSQDVELWEGEVVKELSVBEOIKRNYRYDEDEDE 759

RESULT 2
ABUS2869
ID ABUS2869 standard; Protein; 759 AA.
XX AC ABUS2869;
XX AC AC
XX AC
XX DT 14-APR-2003 (first entry)
XX XX
XX Human transcription factor from DKF2phut1_18c19.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WC200112659-A2.
XX XX
XX PD 22-FEB-2001.
XX XX
XX PF 18-AUG-2000; 2000MO-IB01496.
XX XX
XX PR 18-AUG-1999; 99US-0149499.
XX PR 28-SEP-1999; 99US-0156503.
XX XX
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX XX
XX PI Wiemann S;
XX DR
XX DR WPI: 2001-327840/34.
XX DR N-PSDB; ABX71307.
XX XX

PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX Claim 21; Page 450-451; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a
CC polypeptide described in the disclosure of the invention.
XX
SQ Sequence 759 AA;

Query Match 100.0%; Score 759; DB 22; Length 759;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESSPFNRQWTSLSRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRNTENL 60
Db 1 MESSPFNRQWTSLSRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRNTENL 60
QY 61 SQHFRKGTILTVLKKWENPGLCAESHTDSLNSSTEIHRADHPPEAETVSHAASAKAQDQ 120
Db 61 SQHFRKGTILTVLKKWENPGLCAESHTDSLNSSTEIHRADHPPEAETVSHAASAKAQDQ 120
QY 121 EQIHPRLRLSPPEALVQGRYPHIKDGEDLKDHSTESKMKENCLGESRHEVEKSEISEN 180
Db 121 EQIHPRLRLSPPEALVQGRYPHIKDGEDLKDHSTESKMKENCLGESRHEVEKSEISEN 180
QY 181 TDASKTEKNVPLNRLKMFKEGPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASKTEKNVPLNRLKMFKEGPTQTKILRAQSRSASGRKISENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSEKNESSRNLELRLSETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGGEIKI 300
Db 241 LSSSTFDSEKNESSRNLELRLSETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGGEIKI 300
QY 301 HKMQEKNVPPGPEVCITHEGEKISANENSLAVRSTPAEDDSDSDSQVKSEVQPVHPKP 360
Db 301 HKMQEKNVPPGPEVCITHEGEKISANENSLAVRSTPAEDDSDSDSQVKSEVQPVHPKP 360
QY 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVCEQKTYVPMERLLANQOVPHISCFCRSCYC 420
Db 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVCEQKTYVPMERLLANQOVPHISCFCRSCYC 420
QY 421 NNKLSLGTYSALHGRYICKPHFNOLFYSKGNVDEGFGHRPHKDLWASKNENEILERPAP 480
Db 421 NNKLSLGTYSALHGRYICKPHFNOLFYSKGNVDEGFGHRPHKDLWASKNENEILERPAP 480
QY 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETTKLRIAMPPTTELGS 540
Db 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETTKLRIAMPPTTELGS 540
QY 541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKRSRPTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKRSRPTVAASFQST 600
QY 601 SVKSPKTVSPPIRGWSMSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRGWSMSQSESVGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
QY 721 TONOKSQDVELWEGEVVKELSVBEOIKRNYRYDEDEDE 759
Db 721 TONOKSQDVELWEGEVVKELSVBEOIKRNYRYDEDEDE 759

RESULT 3	
AA039013	
ID	AA039013 standard; Protein; 759 AA.
XX	
AC	AA039013;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 2158.
XX	
KW	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WPI; 2001-442253/47.
XX	
DR	N-PSDB; AA158169.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Example 4; SEQ ID NO 2158; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AA038642-AA042213) with neurotropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilization of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 759 AA;
XX	
Query Match	100.0%; Score 759; DB 22; Length 759;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 759; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MESSPFNRROWTSLSLRVVTAKELSLVNNKSSAIVEIFSKYQKAAEETNNEKRSNTENL	60
DB	1	MESSPFNRROWTSLSLRVVTAKELSLVNNKSSAIVEIFSKYQKAAEETNNEKRSNTENL	60
QY	61	SOHFRKGTTLVLKKWENPOLGAESHTDLSRNSSTIRHRADHPHAEVTSHAASGAKAQ	120
DB	61	SOHFRKGTTLVLKKWENPOLGAESHTDLSRNSSTIRHRADHPHAEVTSHAASGAKAQ	120
QY	121	EQIHRSLRSLRSPPEALVOGRYPHIKDGEDLKDHSTESKVENCLGESRHEVEKSIEN	180
DB	121	EQIHRSLRSLRSPPEALVOGRYPHIKDGEDLKDHSTESKVENCLGESRHEVEKSIEN	180
QY	181	TDASGKIEKYNVPLNRLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDDEIGPGQ	240
DB	181	TDASGKIEKYNVPLNRLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDDEIGPGQ	240
QY	241	LSSTFDSEKNESRRNLELPRISSETIKORMAKYQAAVSKQSSTNTYNELKASGGEIKI	300
DB	241	LSSTFDSEKNESRRNLELPRISSETIKORMAKYQAAVSKQSSTNTYNELKASGGEIKI	300
QY	301	HKWEQENYVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDQVSEVOQPVHPKP	360
DB	301	HKWEQENYVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDQVSEVOQPVHPKP	360
QY	361	LSPDSRASSLSSESSPPKAMKKFOAPARETCVECKTVYPMERLLANQOVFHSICFRCSCYC	420
DB	361	LSPDSRASSLSSESSPPKAMKKFOAPARETCVECKTVYPMERLLANQOVFHSICFRCSCYC	420
QY	421	NNKLSLGTIVASLHGRYICKPHFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIEILRPAQ	480
DB	421	NNKLSLGTIVASLHGRYICKPHFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIEILRPAQ	480
QY	481	LANARETHPSGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGS	540
DB	481	LANARETHPSGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGS	540
QY	541	SGSALBEGIKMSKPKWPPDEISKPEVPEVDVLDLKLKRSLSLKERSPPFTVAASFQST	600
DB	541	SGSALBEGIKMSKPKWPPDEISKPEVPEVDVLDLKLKRSLSLKERSPPFTVAASFQST	600
QY	601	SVKSPXTVSPPIRKGMSESEESVGGVVAERKOVENAKASKKNGVGTWQNKESKG	660
DB	601	SVKSPXTVSPPIRKGMSESEESVGGVVAERKOVENAKASKKNGVGTWQNKESKG	660
QY	661	ETGKRSGHSELENNENLVENGADSDDDNDFLQKQSPQEPKSLNWSFVDTTAAEFT	720
DB	661	ETGKRSGHSELENNENLVENGADSDDDNDFLQKQSPQEPKSLNWSFVDTTAAEFT	720
QY	721	TQNKSDQVELWEGEVVVELSVEEQIKRNRYYDEDEDEE	759
DB	721	TQNKSDQVELWEGEVVVELSVEEQIKRNRYYDEDEDEE	759
RESULT 4			
AAB67701			
ID	AAB67701	standard; Protein; 759 AA.	
XX			
AC	AAB67701;		
XX			
DT	11-JUN-2001	(first entry)	
XX			
DE	Human EPLIN (epithelial protein lost in neoplasm)-beta isoform.		
KW	Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;		
KW	EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;		
XX	gene therapy; cancer.		
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FT	Misc-difference 278		
FT	/note= "Val encoded by GCG"		
FT	Misc-difference 344		

FT Misc-difference 495 /note= "Arg encoded by CCAGGT"
 FT /note= "Asp encoded by AAT"
 FT Misc-difference 511 /note= "Lys encoded by AGG"
 FT Misc-difference 622 /note= "Ser encoded by AAT"
 FT Misc-difference 658 /note= "Ser encoded by TTT"
 FT Misc-difference 679 /note= "Ser encoded by TTT"
 FT /note= "Ile encoded by TTT"
 FT Misc-difference 698 /note= "Ser encoded by TTT"
 FT Misc-difference 704 /note= "Ser encoded by TTT"
 FT
 PN WO200118019-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000WO-US24689.
 XX
 PR 08-SEP-1999; 99US-0153024.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Chang DD, Maul RS;
 XX
 XX WPI: 2001-244555/25.
 DR N-PSDB; AAF55697.
 XX
 XX New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
 PT prognostic and therapeutic applications over the course of cell
 PT proliferative disorders associated with EPLIN -
 XX
 PS Claim 1; Page 44; 59pp; English.
 XX
 CC The present sequence represents a human EPLIN (epithelial protein lost in
 CC neoplasm)-beta isoform. The specification also describes EPLIN-alpha.
 CC EPLIN is a tumour suppressor protein, whose expression is altered in
 CC multiple common human tumour types. EPLIN nucleic acids and proteins are
 CC used in screening assays to detect molecules that specifically bind to
 CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
 CC as agonist or antagonist of EPLIN, in particular molecules that affect
 CC cell proliferation. Thus the assays are useful for screening molecules
 CC with potential utility as anticancer drugs or lead compounds for drug
 CC development. EPLIN nucleic acids, proteins are useful for detecting a
 CC cell proliferative disorder in a subject. EPLIN polynucleotides are
 CC useful in gene therapy techniques. EPLIN is useful as a marker that
 CC can be diagnostically, prognostically and therapeutically used over
 CC the course of a cell proliferative disorder associated with EPLIN.
 XX
 SQ Sequence 759 AA;
 Query Match 89,3%; Score 678; DB 22; Length 759;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MESSPFNRRQWTSLSRVTAKEISLVNKNKSSAIVEIFSKYQAAETNMKRSNTENL 60
 Db 1 MESSPFNRRQWTSLSRVTAKEISLVNKNKSSAIVEIFSKYQAAETNMKRSNTENL 60
 QY 61 SQHFRKGTTLVLKKNWPNGLGAESHTDLSRNSSTEIRHRADHPHAEVTSAAAGAKADQ 120
 Db 61 SQHFRKGTTLVLKKNWPNGLGAESHTDLSRNSSTEIRHRADHPHAEVTSAAAGAKADQ 120
 QY 121 EQQHPRSLRSLRSPPEALVQGRPHIKDGEDLKHSTESKXVENCNLCGESRHEVEKSEISEN 180
 Db 121 EQQHPRSLRSLRSPPEALVQGRPHIKDGEDLKHSTESKXVENCNLCGESRHEVEKSEISEN 180
 QY 181 TDASGKIEKNVPLNRLKXWFKFGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
 Db 181 TDASGKIEKNVPLNRLKXWFKFGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240

QY 241 LSSSTFDSEKXNRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
 Db 241 LSSSTFDSEKXNRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
 QY 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSDRDSDQVSEVQVPHPKP 360
 Db 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSDRDSDQVSEVQVPHPKP 360
 QY 361 LSPDRASSLSSESSPPKAMKKFOAPARETCVECOQTVYPMERLLANQQVFHISCFCRCSYC 420
 Db 361 LSPDRASSLSSESSPPKAMKKFOAPARETCVECOQTVYPMERLLANQQVFHISCFCRCSYC 420
 QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNVDGFGHPRHDKLWASKNENESILERPAAQ 480
 Db 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNVDGFGHPRHDKLWASKNENESILERPAAQ 480
 QY 481 LANARETHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAMPPTTELGS 540
 Db 481 LANARETHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAMPPTTELGS 540
 QY 541 SGSALBEGIKMSPKWPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPFTVAASFQST 600
 Db 541 SGSALBEGIKMSPKWPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPFTVAASFQST 600
 QY 601 SVKSPKTVSPPIRKGWMSGEQSEESYVGRVARKQVENAKASKKNGVNGKTTWQNKESKG 660
 Db 601 SVKSPKTVSPPIRKGWMSGEQSEESYVGRVARKQVENAKASKKNGVNGKTTWQNKESKG 660
 QY 661 ETGKRSKEGHSLEMEENEN 678
 Db 661 ETGKRSKEGHSLEMEENEN 678
 RESULT 5
 AAB42934
 ID AAB42934 standard; Protein: 760 AA.
 XX
 AC AAB42934;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2698 polypeptide sequence SEQ ID NO:5396.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; aniprositic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.

[illegible]

Db	481	QLANARETPHSPGVDEAPIAKVGVLAAWEAKASSQCEKEDKPAETKKLRIAWPPPTELG	540
Qy	540	SSGSALLEGIKMSXPKNPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQS	599
Db	541	SSGSALLEGIKMSXPKNPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQS	600
Qy	600	TSVSKSPKTVSPPIEKGMSSQSEESVCGRVAERKQVENAKSKNGNVGKTTWONKESK	659
Db	601	TSVSKSPKTVSPPIIRKGMSSQSEESVCGRVAERKQVENAKSKNGNVGKTTWONKESK	660
Qy	660	GETGKRSGKEGHSLEMEENLVENGADSDDEDDNSFLKQOSPQPKSLNWSFVDNTFAEEF	719
Db	661	GETGKRSGKEGHSLEMEENLVENGADSDDEDDNSFLKQOSPQPKSLNWSFVDNTFAEEF	720
Qy	720	TTQNKQSQDVELWEGEVYVVKLSVEEQIKRNRYYDEDEEE	759
Db	721	TTQNKQSQDVELWEGEVYVVKLSVEEQIKRNRYYDEDEEE	760
RESULT 6			
AAB95477	ID	AAB95477 standard; Protein; 457 AA.	
XX	AC	AAB95477;	
XX	DT	26-JUN-2001 (first entry)	
XX	DE	Human protein sequence SEQ ID NO:17592.	
XX	KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
XX	OS	Homo sapiens.	
XX	PN	EP1074617-A2.	
XX	PD	07-FEB-2001.	
XX	PF	28-JUL-2000; 2000EP-0116126.	
XX	PR	29-JUL-1999; 99JP-0248036.	
XX	PR	27-AUG-1999; 99JP-0300253.	
XX	PR	11-JAN-2000; 2000JP-0118776.	
XX	PR	02-MAY-2000; 2000JP-0183767.	
XX	PR	09-JUN-2000; 2000JP-0241899.	
XX	PA	(HELI-) HELIX RES INST.	
XX	PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
XX	PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX	DR	WPI; 2001-318749/34.	
XX	PT	Primer sets for synthesizing polynucleotides, particularly the 5602	
XX	PT	full-length cDNAs defined in the specification, and for the detection	
XX	PT	and/or diagnosis of the abnormality of the proteins encoded by the	
XX	PT	full-length cDNAs -	
XX	PS	Claim 8; SEQ ID 17992; 2537pp + CD ROM; English.	
XX	CC	The present invention describes primer sets for synthesizing 5602	
XX	CC	full-length cDNAs defined in the specification. Where a primer set	
XX	CC	comprises: (a) an oligo-dr primer and an oligonucleotide complementary	
XX	CC	to the complementary strand of a polynucleotide which comprises one of	
XX	CC	the 5602 nucleotide sequences defined in the specification, where the	
XX	CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
XX	CC	of an oligonucleotide comprising a sequence complementary to the	
XX	CC	complementary strand of a polynucleotide which comprises a 5'-end	
XX	CC	sequence and an oligonucleotide comprising a sequence complementary to a	
XX	CC	polynucleotide which comprises a 3'-end sequence, where the	
XX	CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
XX	CC	the 5'-end sequence/3'-end sequence is selected from those defined in	
XX	CC	the specification. The primer sets can be used in antisense therapy and	

XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT Misc-difference 118 /note= "Val encoded by GCG"
XX FT Misc-difference 336 /note= "Asp encoded by AAT"
XX FT Misc-difference 352 /note= "Lys encoded by AGG"
XX FT Misc-difference 463 /note= "Ser encoded by AAT"
XX FT Misc-difference 499 /note= "Ser encoded by TTT"
XX FT Misc-difference 520 /note= "Ile encoded by TTT"
XX FT Misc-difference 539 /note= "Ser encoded by TTT"
XX FT Misc-difference 545 /note= "Ser encoded by TTT"
XX PN WO200118019-A1.
XX PD 15-MAR-2001.
XX PF 08-SEP-2000; 2000WO-US24689.
XX PR 08-SEP-1999; 99US-0153024.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Chang DD, Maul RS;
XX DR WPI; 2001-244555/25.
XX DR N-PSDB; AAF5696.
XX PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
XX PT prognostic and therapeutic applications over the course of cell
XX PT proliferative disorders associated with EPLIN -
XX PS Claim 1; Page 43; 59pp; English.
XX CC The present sequence represents a human EPLIN (epithelial protein lost in
XX CC neoplasm)-alpha isoform. The specification also describes EPLIN-beta.
XX CC EPLIN is a tumour suppressor protein, whose expression is altered in
XX CC multiple common human tumour types. EPLIN nucleic acids and proteins are
XX CC used in screening assays to detect molecules that specifically bind to
XX CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
XX CC as agonist or antagonist of EPLIN in particular molecules that affect
XX CC cell proliferation. Thus the assays are useful for screening molecules
XX CC with potential utility as anticancer drugs or lead compounds for drug
XX CC development. EPLIN nucleic acids, proteins are useful for detecting a
XX CC cell proliferative disorder in a subject. EPLIN polynucleotides are
XX CC useful in gene therapy techniques. EPLIN is useful as a marker that
XX CC can be diagnostically, prognostically and therapeutically used over
XX CC the course of a cell proliferative disorder associated with EPLIN.
XX SQ Sequence 600 AA;
XX Query Match 44.0%; Score 334; DB 22; Length 600;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 345 DSQKSEVQVPHKPLSPDSSSLSESSPPKAMKFFQAPARETCVECKTVYPMERLL 404
DB 186 DSQKSEVQVPHKPLSPDSSSLSESSPPKAMKFFQAPARETCVECKTVYPMERLL 245
OY 405 ANQOVFHSRCYCNKLSLGTYSALHGRYICKPHFNQLFKSGNYDGFGRPHKOL 464
DB 246 ANQOVFHSRCYCNKLSLGTYSALHGRYICKPHFNQLFKSGNYDGFGRPHKOL 305
OY 465 WASKNENEILERPAQLANARETTPHSPGVEDAPIAKVGVLAAASWEAKASSQOQEKDPAE 524

DB 306 WASKNENEILERPAQLANARETTPHSPGVEDAPIAKVGVLAAASWEAKASSQOQEKDPAE 365
OY 525 TKKLRIAMPPPTLGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKKLRRSSSL 584
DB 366 TKKLRIAMPPPTLGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKKLRRSSSL 425
OY 585 KERSRPTVAASFOSTSVKSPKTVSPPIRKGMWSQESESVGGRVAERKOVENAKASKK 644
DB 426 KERSRPTVAASFOSTSVKSPKTVSPPIRKGMWSQESESVGGRVAERKOVENAKASKK 485
OY 645 NGNVGKTTWQNKESKGETGKRSKEGHSLENNEN 678
DB 486 NGNVGKTTWQNKESKGETGKRSKEGHSLENNEN 519
RESULT 9
AAB34206
ID AAB34206 standard; Protein; 299 AA.
XX AC AAB34206;
XX DT 26-JAN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:174.
XX KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
XX KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
XX KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;
XX KW infection; ocular disorder; wound healing; skin aging; food additive;
XX KW preservative.
XX OS Homo sapiens.
XX PN WO200056755-A1.
XX XX 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06830.
XX PR 19-MAR-1999; 99US-0125361.
XX PR 10-DEC-1999; 99US-0169910.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SW, Komatsoulis G;
XX DR WPI; 2000-587661/55.
XX PT New isolated nucleic acid molecules encoding 49 human secreted proteins
XX PT used for preventing, treating or ameliorating medical conditions, for
XX PT diagnosing pathological conditions or as food additives or
XX PS preservatives -
XX PS Disclosure; Page 412-413; 419pp; English.
XX CC The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX CC AAB34216 represent human secreted polypeptide sequences and proteins
XX CC homologous to them, which are given in the exemplification of the present
XX CC invention. Human secreted proteins have activities based on the tissue
XX CC and cells the genes are expressed in. Examples of activities include:
XX CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
XX CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX CC and vulnerary. The polynucleotides and polypeptides can be used to
XX CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX CC in diagnosing a pathological condition or susceptibility to a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or

CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
CC in the exemplification of the present invention.

XX Sequence 299 AA;
SQ
Query Match 39.4%; Score 299; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.6e-286; Length 299;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 454 EGFGRPHKDLWASKNEEILERPALANARETPHSPGVEDAPIAKVGVLAASMEAKAS 513
Db 1 EGFGRPHKDLWASKNEEILERPALANARETPHSPGVEDAPIAKVGVLAASMEAKAS 60
QY 514 SQQEKDKPAETKKLRITAWPPPTLGGSSALERGIKMSKPKWPPDEISKPEVPEDVDL 573
Db 61 SQQEKDKPAETKKLRITAWPPPTLGGSSALERGIKMSKPKWPPDEISKPEVPEDVDL 120
QY 574 DLKKLRSSSLKERSRPTVAASFOSTSVKSPKTVSPPIRGWMSMSQSESVGGRVAER 633
Db 121 DLKKLRSSSLKERSRPTVAASFOSTSVKSPKTVSPPIRGWMSMSQSESVGGRVAER 180
QY 634 KOVENAKASKKNGVNGKTTWQNKESKGETGKRSGEHSLEMEENLVENGADSDDDNSF 693
Db 181 KOVENAKASKKNGVNGKTTWQNKESKGETGKRSGEHSLEMEENLVENGADSDDDNSF 240
QY 694 LKQSPQEPKSLNWSFVDNTFAEFTTQNKQODVELWEGEVVVKLSVEEQIKRNY 752
Db 241 LKQSPQEPKSLNWSFVDNTFAEFTTQNKQODVELWEGEVVVKLSVEEQIKRNY 299

RESULT 10
ABBS6420
ID ABB56420 standard; Protein; 301 AA.
XX
AC ABB56420;
XX
DT 21-FEB-2002 (first entry)
XX
DE Human cancer suppressor protein PP624.
XX
KW Human; cancer suppressor; disease; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 21 /label= unknown
FT /note= "Encoded by RCA"
FT Misc-difference 40 /label= unknown
FT /note= "Encoded by TRC"
FT Misc-difference 43 /label= unknown
FT /note= "Encoded by AKC"
XX
XX
PN CN1313297-A.
XX
PD 19-SEP-2001.
XX
PF 09-MAR-2000; 2000CN-0111948.
XX
PR 09-MAR-2000; 2000CN-0111948.
XX
PA (SHAN-) SHANGHAI INST ONCOLOGY.

XX Gu J, Yang S;
XX WPI; 2002-042185/06.
XX N-PSDB; ABI98973.
XX Human protein able to suppress growth of cancer cells and its coding
XX sequence -
XX Claim 1; Page 20 Disclosure; 37pp; Chinese.
XX
CC The invention relates to novel human proteins (ABB56417-ABB56425) with
CC cancer suppressing function, the encoding polynucleotides
CC (ABI98970-ABI98978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
XX
SQ Sequence 301 AA;
Query Match 31.6%; Score 240; DB 23; Length 301;
Best Local Similarity 100.0%; Pred. No. 1e-227;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 422 NKLSLGTVASLHGRYCKPHFNQLFKSKGNVDEGFGRPHKDLWASKNEEILERPQOL 481
Db 44 NKLSLGTVASLHGRYCKPHFNQLFKSKGNVDEGFGRPHKDLWASKNEEILERPQOL 103
QY 482 ANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRAMPPTTELSS 541
Db 104 ANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRAMPPTTELSS 163
QY 542 GSALKEGKMSKPKWPPDEISKPEVPEDVDLKLKRRSSSLKERSRPTVAASFQSTS 601
Db 164 GSALKEGKMSKPKWPPDEISKPEVPEDVDLKLKRRSSSLKERSRPTVAASFQSTS 223
QY 602 VKSPKTVSPPIRGWMSMSQSESVGGRVAERKOVENAKASKKNGVNGKTTWQNKESKGE 661
Db 224 VKSPKTVSPPIRGWMSMSQSESVGGRVAERKOVENAKASKKNGVNGKTTWQNKESKGE 283

RESULT 11
AAU10979
ID AAU10979 standard; Protein; 457 AA.
XX
AC AAU10979;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human sterol regulatory element binding protein 3.
XX
KW Human; sterol regulatory element binding protein 3; hSRBP-3;
KW hypocholesterolemia.
XX
OS Homo sapiens.
XX
PN CN1309182-A.
XX
PD 22-AUG-2001.
XX
PF 17-FEB-2000; 2000CN-0111698.
XX
PR 17-FEB-2000; 2000CN-0111698.
XX
PA (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.
XX
PI Li Y, Xu S, Ren S;
XX WPI; 2002-011822/02.
XX N-PSDB; AAS18588.
XX
XX Cholesterol regulatory factor binding protein and its coding sequence -

PS Claim 2; Page 14; 27pp; Chinese.

XX The invention relates to a novel human sterol regulatory element binding

CC protein 3 (hsREBP-3) expressed in human normal hypothalamic tissue and

CC its coding sequence. Also described is the process for preparing the

CC protein and nucleic acid sequence, and the method for detecting hsREBP-3

CC nucleic acid sequence and polypeptides. The present sequence represents

CC the amino acid sequence of human hsREBP-3 as described in the invention.

XX

SQ Sequence 457 AA;

Query Match 31.4%; Score 238; DB 23; Length 457;

Best Local Similarity 100.0%; Pred. No. 1.4e-225; Indels 0; Gaps 0;

Matches 238; Conservative 0; Mismatches 0;

QY 522 PATETKRLIAMPPTTELSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRS 581

Db 220 PATETKRLIAMPPTTELSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRS 279

QY 582 SSLKERSPPFTVAASFQSTSVKSPKTVPPPIRKGWSMSEQSEESVGGVRAERKQVENAKA 641

Db 280 SSLKERSPPFTVAASFQSTSVKSPKTVPPPIRKGWSMSEQSEESVGGVRAERKQVENAKA 339

QY 642 SKKNGNVGKTTWQNKESKGETGKRSKEGHSLEMENLVENGADSDDDNSFLKQSQPQE 701

Db 340 SKKNGNVGKTTWQNKESKGETGKRSKEGHSLEMENLVENGADSDDDNSFLKQSQPQE 399

QY 702 PKSLNWSFVNTFAEFTTQNKSQDVLEWGEVVKLSVEEQIKRNYRYDEDEEE 759

Db 400 PKSLNWSFVNTFAEFTTQNKSQDVLEWGEVVKLSVEEQIKRNYRYDEDEEE 457

RESULT 12

AAB51883

ID AAB51883 standard; Protein; 232 AA.

XX

AC AAB51883;

XX

DT 16-FEB-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.

XX

KW Human; secreted protein; immunosuppressive; antiarthritic; anirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorders; cancer; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; nervous system disorder;

KW Alzheimer's disease; ocular disorder; wound healing; skin aging.

XX

OS Homo sapiens.

XX

PN WO200061626-A1.

XX

PD 19-OCT-2000.

XX

EF 06-APR-2000; 2000WO-US09066.

XX

PR 09-APR-1999; 99US-0128698.

PR 20-JAN-2000; 2000US-0176926.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE)/ ROSEN C A.

XX

XX Rosen CA, Ruben SM, Komatsoulis G;

XX

XX WPI; 2000-619227/59.

DR N-PSDB; AAC93483.

XX

XX New nucleic acid molecules encoding 49 human secreted proteins for

PT diagnosing, preventing or ameliorating medical conditions and used for

PT food additives or preservatives -

XX

PS Disclosure; Page 22; 516pp; English.

XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding

CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -

CC AAB51927 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and

CC proteins have activities dependent on the tissues and cells in which they

CC are expressed. Examples of their activities include immunosuppressive;

CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;

CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;

CC virucide; fungicide; ophthalmological; and vulnerary. The secreted

CC proteins, polynucleotides, antagonists and agonists may be useful in

CC treating, preventing and/or diagnosing diseases and disorders such as

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

CC carbohydrate, vitamins, minerals, cofactors and other nutritional

CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are

CC used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX

SQ Sequence 232 AA;

Query Match 29.9%; Score 227; DB 21; Length 232;

Best Local Similarity 100.0%; Pred. No. 5.4e-215; Indels 0; Gaps 0;

Matches 227; Conservative 0; Mismatches 0;

QY 473 EILERPAQLANARETPHSPGVEDAPIAKGVGLAAWEAKASSQOQKEDKPAETKRLIAW 532

Db 1 EILERPAQLANARETPHSPGVEDAPIAKGVGLAAWEAKASSQOQKEDKPAETKRLIAW 60

QY 533 PPTELGGSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRSSSLKRSRPT 592

Db 61 PPTELGGSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRSSSLKRSRPT 120

QY 593 VAASFQSTSVKSPKTVSPPIRKGWSMSQSEESVGGVRAERKQVENAKSKNGNVGKTT 652

Db 121 VAASFQSTSVKSPKTVSPPIRKGWSMSQSEESVGGVRAERKQVENAKSKNGNVGKTT 180

QY 653 WQNKESKGETGKRSKEGHSLEMENLVENGADSDDDNSFLKQSQP 699

Db 181 WQNKESKGETGKRSKEGHSLEMENLVENGADSDDDNSFLKQSQP 227

RESULT 13

AAB54159

ID AAB54159 standard; Protein; 243 AA.

XX

AC AAB54159;

XX

DT 09-MAR-2001 (first entry)

XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:611.

XX

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

KW detection; diagnosis; identification; cytostatic; neuroprotective;

KW neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological;

KW antiinflammatory; cardiant; gene therapy; chromosome mapping;

KW linkage analysis; tissue identification; tissue typing; forensic;

KW neural; immune system; muscular; reproductive; gastrointestinal;

KW pulmonary; cardiovascular; renal; proliferative.

XX

OS Homo sapiens.

XX

PN WO200055320-A1.


```
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05989.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI; 2000-579444/54.
DR
XX N-PSDB; AAC98924.
DR
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 11; Page 1050-1051; 1379pp; English.
PS
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontoxic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiac and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridization probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 243 AA;
SQ
Query Match 24.2%; Score 184; DB 21; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.4e-172;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 EGEKISANENSLAVRSTPAEDDSDSDSOVKSEVQVPHKPLSPDSDRASSLSSESSPKAMK 380
Db 1 EGEKISANENSLAVRSTPAEDDSDSDSOVKSEVQVPHKPLSPDSDRASSLSSESSPKAMK 60
QY 381 KFOAPARETCVEQKTYVPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRITYCKP 440
Db 61 KFOAPARETCVEQKTYVPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRITYCKP 120
QY 441 HFNLQFKSKNGYDEGCHRPKOLWASKNENEFILRPAQLANARETPHSPGVEDAPIAK 500
Db 121 HFNLQFKSKNGYDEGCHRPKOLWASKNENEFILRPAQLANARETPHSPGVEDAPIAK 180
QY 501 VGVLL 504
Db 181 VGVLL 184
RESULT 14
AAV73916
ID AAV73916 standard; Protein; 235 AA.
XX
XX AAV73916;
AC
XX
XX 14-MAR-2000 (first entry)
DT
```

```
XX Human prostate tumor EST fragment derived protein #103.
DE
XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX treatment.
KW
XX Homo sapiens.
OS
XX DE19820190-A1.
PN
XX 04-NOV-1999.
PD
XX
XX 28-APR-1998; 98DE-1020190.
PF
XX
XX 28-APR-1998; 98DE-1020190.
PR
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
PI
XX WPI; 1999-621386/54.
DR
XX N-PSDB; AAZ52892.
DR
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
XX Claim 23; Page 351; 502pp; German.
PS
XX
XX This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAY73914-Y74252
CC represent protein fragments encoded by the human pancreatic tumor cDNA
CC library derived expressed sequence tag (EST) sequences represented in
CC AAZ52858-Z53014.
XX
XX Sequence 235 AA;
SQ
Query Match 22.4%; Score 170; DB 20; Length 235;
Best Local Similarity 100.0%; Pred. No. 9e-159;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MESSPFNRQRQWTSLSRVTAKELSLVNKNKSSAIVEIFSFKYKAAEETNMEKRSNTENL 60
Db 10 MESSPFNRQRQWTSLSRVTAKELSLVNKNKSSAIVEIFSFKYKAAEETNMEKRSNTENL 69
QY 61 SQHFRKGLTTLVKKXWNPGLGAESHTDSLNSSTEIRHRAHPPAEVTSAAASGAKAQD 120
Db 70 SQHFRKGLTTLVKKXWNPGLGAESHTDSLNSSTEIRHRAHPPAEVTSAAASGAKAQD 129
QY 121 EQIHPRLSPPEALVQGRYPHIKDGEDLKDHSTESKKNVCLGESRH 170
Db 130 EQIHPRLSPPEALVQGRYPHIKDGEDLKDHSTESKKNVCLGESRH 179
RESULT 15
ABG24090
ID ABG24090 standard; Protein; 337 AA.
XX
XX AC ABG24090;
AC
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #24081.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
PD
```

XX 30-MAR-2001; 2001WO-US08631.
PF
XX
PR 31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS88277.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 54449; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 337 AA;
SQ

Query Match 20.4%; Score 155; DB 22; Length 337;
Best Local Similarity 100.0%; Pred.No. 7.8e-144;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NTELSQHFRKGTLLVLKXKXWENPGLGSHSTDSLRNSSTIRHRADHPPAEVTSAAAG 115
Db 2 NTELSQHFRKGTLLVLKXKXWENPGLGSHSTDSLRNSSTIRHRADHPPAEVTSAAAG 61
QY 116 AKADQEQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKXVNCJGESRHEVEKS 175
Db 62 AKADQEQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKXVNCJGESRHEVEKS 121
QY 176 EISENTDASGKIEKYNVPLNKLKMMFEKGEPTQTK 210
Db 122 EISENTDASGKIEKYNVPLNKLKMMFEKGEPTQTK 156

Search completed: January 6, 2004, 09:51:42
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:50:00 ; Search time 21 Seconds
(without alignments)
3475.811 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPNNRQWTSLSLRVTA.....LSVESQIKRNYRDEDEEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.1	122	2 E90243	hypothetical prote
2	8	1.1	231	2 T02585	hypothetical prote
3	8	1.1	234	2 G72758	hypothetical prote
4	8	1.1	388	2 D72740	hypothetical prote
5	8	1.1	403	2 S71236	probable N-acetyl
6	8	1.1	405	2 T27971	hypothetical prote
7	8	1.1	413	2 A84620	similar to hookles
8	8	1.1	415	2 S00543	site-specific DNA-
9	8	1.1	466	2 E95262	L-seryl-tRNAse
10	8	1.1	642	1 JF0079	LIM protein kinase
11	8	1.1	784	2 H83158	Fe(III) dicitrate
12	8	1.1	792	2 F70388	grase B - Aquifex
13	8	1.1	1038	2 I38935	bone morphogenetic
14	8	1.1	1038	2 J05527	probable membrane
15	8	1.1	2295	2 B71621	hypothetical prote
16	8	1.1	2484	2 T26216	hypothetical prote
17	8	1.1	2607	2 A56388	dopamine- and cAMP
18	7	0.9	49	2 AB2271	periplasmic mercur
19	7	0.9	64	2 G64010	hypothetical prote
20	7	0.9	85	2 S20505	nodulin-24 - soybe
21	7	0.9	92	2 T31048	hypothetical prote
22	7	0.9	93	2 H72542	13K transport prot
23	7	0.9	101	2 S45096	13K transport prot
24	7	0.9	118	2 S45099	hypothetical prote
25	7	0.9	118	2 T49168	hypothetical prote
26	7	0.9	121	2 C90510	hypothetical prote
27	7	0.9	124	2 T37334	hypothetical prote
28	7	0.9	134	2 T10068	Probable hexose tr
29	7	0.9	132	2	

30	7	0.9	135	2 B38253	kappa-enhancer-bin
31	7	0.9	139	2 S69458	hypothetical prote
32	7	0.9	139	2 B96500	hypothetical prote
33	7	0.9	141	2 T09581	probable high mobi
34	7	0.9	149	2 S39556	high mobility grou
35	7	0.9	153	2 A71893	hypothetical prote
36	7	0.9	159	2 C49773	ecdysone-dependent
37	7	0.9	159	2 A72104	hypothetical prote
38	7	0.9	159	2 B86518	hypothetical prote
39	7	0.9	160	2 AC2077	hypothetical prote
40	7	0.9	169	2 AH1032	hypothetical prote
41	7	0.9	177	2 A75446	hypothetical prote
42	7	0.9	178	2 AH1394	hypothetical prote
43	7	0.9	178	2 S26481	B. subtilis RNA po
44	7	0.9	181	2 F82750	calcium-binding pr
45	7	0.9	182	2 B32256	outer membrane pro
					hypothetical 20K p

ALIGNMENTS

RESULT 1
E90243
hypothetical protein SSO0922 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: E90243
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
arrett, R.A.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90243
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KUR>
A:Cross-references: GB:AE006641; NID:g13814104; PIDN:AAK41204.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO0922

Query Match 1.1%; Score 8; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 IEKYNVPL 194
Db 37 IEKYNVPL 44
RESULT 2
T02585
hypothetical protein At2g39170 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T16B24.19
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Feb-2001
C:Accession: T02585; A84814
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02585
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-231 <ROU>
A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402687
A:Experimental source: Cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: A84814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <STO>
A:Cross-references: GB:AE002093; NID:G3402687; PIDN:AAC28990.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g39170; T16B24.19
A:Map position: 2

Query Match 1.1%; Score 8; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 VGGRAER 633
DB 223 VGGRAER 230

RESULT 3
G72758
hypothetical protein APE0060 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72758
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: G72758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <KAW>
A:Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA78969.1; PID:d1042745; PID:G5103388
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0060

Query Match 1.1%; Score 8; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 ESUGRVA 631
DB 199 ESUGRVA 206

RESULT 4
D72740
hypothetical protein APE0453 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: D72740
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: D72740
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <KAW>
A:Cross-references: DDBJ:AP000059; NID:G5103911; PIDN:BAA79416.1; PID:G5104100
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0453
C:Superfamily: Aeropyrum pernix hypothetical protein APE0453

Query Match 1.1%; Score 8; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 VLAASMEA 510

Db 97 VLAASMEA 104

RESULT 5
S71236
probable N-acetyltransferase (EC 2.3.1.-) hookless 1 - Arabidopsis thaliana
N:Alternate names: ethylene response protein; protein F19F18.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 11-Jun-1999
C:Accession: S71236; T04715
R:Jehman, A.L.; Black, R.; Ecker, J.R.
submitted to the EMBL Data Library, March 1996
A:Description: Hookless1, an ethylene response gene, is required for differential cell elongation in Arabidopsis thaliana
A:Reference number: S71236
A:Accession: S71236
A:Molecule type: mRNA
A:Residues: 1-403 <LEH>
A:Cross-references: EMBL:U50399; NID:G1277089; PID:G1277090
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15382
A:Accession: T04715
A:Molecule type: DNA
A:Residues: 1-403 <BEV>
A:Cross-references: EMBL:AL035605
A:Experimental source: cultivar Columbia; BAC clone F19F18
C:Genetics:
A:Gene: hookless1; HLS1
A:Map position: 4
A:Introns: 54/3; 113/2
A:Note: F19F18.70
C:Function:
A:Description: involved in differential cell elongation in the hypocotyl
C:Keywords: acyltransferase

Query Match 1.1%; Score 8; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428
DB 215 NNKLSLGT 222

RESULT 6
T27971
hypothetical protein ZK675.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T27971
R:Sims, M.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z20448
A:Accession: T27971
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-405 <WIL>
A:Cross-references: EMBL:Z46812; PIDN:CAA86845.1; GSPDB:GN00020; CESP:ZK675.3
A:Experimental source: clone ZK675
C:Genetics:
A:Gene: CESP:ZK675.3
A:Map position: 2
A:Introns: 43/2; 84/2; 110/3; 233/3; 264/2; 335/3; 372/1
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK675.3

Query Match 1.1%; Score 8; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 SKNENEI 474
DB 254 SKNENEI 261

```

A;Residues: 1-466 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK64663.1; PID:g14523060; GSPDB:GN00165
A;Experimental source: Strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure,
heault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: selsa
A;Genome: plasmid
C;Superfamily: Escherichia coli seryl-tRNA(Sec) selenium transferase selsa

Query Match 1.1% Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 GSGGSALE 546
|||||
Db 411 GSGGSALE 418

RESULT 10
JP0079
LIM protein kinase (EC 2.7.1.1) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JP0079
R;Ohashi, K.; Toshima, J.; Tajinda, K.; Nakamura, T.; Mizuno, K.
J. Biochem. 116, 636-642, 1994
A;Title: Molecular cloning of a chicken lung cDNA encoding a novel protein kinase with N
A;Reference number: JP0079; MUID:95155277; PMID:7852284
A;Accession: JP0079
A;Molecule type: mRNA
A;Residues: 1-642 <OHA>
A;Cross-references: DDBJ:D26310; NID:g1402513; PIDN:BAA05372.1; PID:g643086
C;Comment: This protein plays an important role in the regulation of diverse cell function
C;Superfamily: LIM protein kinase; LIM metal-binding repeat homology; protein kinase hom
C;Keywords: ATP; duplication; phosphotransferase; serine/threonine/tyrosine-specific pro
F.12-63/Domain: LIM metal-binding repeat homology <LIM>
F.72-124/Domain: LIM metal-binding repeat homology <LIM2>
F.329-608/Domain: protein kinase homology <KIN>
F.337-345/Region: protein kinase ATP-binding motif
F.496-506/Region: basic

Query Match 1.1% Score 8; DB 1; Length 642;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 HISCFCRS 418
|||||
Db 34 HISCFCRS 41

RESULT 11
H83158
Fe(III) dicitrate transport protein FecA PA3901 [imported] - Pseudomonas aeruginosa (str
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83158
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83158
A;Status: preliminary
A;Molecule type: DNA

```

```

RESULT 7
A84620
Similar to hookless1 (HLS1) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84620
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanNaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84620
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-413 <STO>
A;Cross-references: GB:AE002093; NID:g3169179; PIDN:AAC17822.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g23060
A;Map position: 2

Query Match 1.1% Score 8; DB 2; Length 413;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428
|||||
Db 225 NNKLSLGT 232

RESULT 8
S00543
site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) DdeI - Desulfovibrio
A;Alternate names: modification methylase DdeI
C;Species: Desulfovibrio desulfuricans
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 05-May-2000
C;Accession: S00543
R;Snyder, L.A.; Slacko, B.; Moran, L.; O'Donnell, K.H.; Brooks, J.E.
Nucleic Acids Res. 15, 8249-8286, 1987
A;Title: Nucleotide sequence of the DdeI restriction-modification system and characteriz
A;Reference number: S00542; MUID:88040453; PMID:2823226
A;Accession: S00543
A;Molecule type: DNA
A;Residues: 1-415 <SNZ>
A;Cross-references: EMBL:Y00449; NID:g40793; PIDN:CAA68505.1; PID:g40795
C;Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 1.1% Score 8; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 TVAASTQS 599
|||||
Db 296 TVAASTQS 303

RESULT 9
E95262
-seryl-tRNAsec selenium transferase (EC 2.9.1.1) [imported] - Sinorhizobium meliloti (s
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 03-Jun-2002
C;Accession: E95262
R;Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
roc. Natl. Acad. Sci. U.S.A. 98, 9883-9886, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: E95262
A;Status: preliminary
A;Molecule type: DNA

```

A;Residues: 1-784 <STO>
A;Cross-references: GB:AE004807; GB:AE004091; NID:g950076; PIDN:AAG07288.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: fcaA; PA3901
C;Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to

Query Match 1.1%; Score 8; DB 2; Length 784;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 LGESRHEV 172
Db 440 LGESRHEV 447

RESULT 12
F70388
Gyrase B - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C;Accession: F70388
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: F70388
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-792 <AQF>
A;Cross-references: GB:AE000719; NID:g2983517; PIDN:AAC07098.1; PID:g2983522; GB:AE00065
A;Experimental source: strain VFS
C;Genetics:
A;Gene: gyrB
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 1.1%; Score 8; DB 2; Length 792;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 DLDLKKLR 579
Db 486 DLDLKKLR 493

RESULT 13
I38935
bone morphogenetic protein receptor II precursor - human
N;Alternate names: activin receptor-like kinase type II; bone morphogenetic protein 4 re
N;Contains: protein kinase (EC 2.7.1.37)
C;Species: Homo sapiens (man)
C;Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 24-Sep-1999
C;Accession: I38935; I55438; I37209
R;Kawabata, M.; Chytil, A.; Mosses, H.L.
J. Biol. Chem. 270, 5625-5630, 1995
A;Title: Cloning of a novel type II serine/threonine kinase receptor through interaction
A;Reference number: A55947; MUID:95197572; PMID:7890683
A;Accession: I38935
A;Molecule type: mRNA
A;Residues: 1-1038 <KAW>
A;Cross-references: ENBL:U20165; NID:g704361; PIDN:AAC50105.1; PID:g704362
R;Nonno, T.; Ishikawa, T.; Saito, T.; Hosokawa, K.; Noji, S.; Wolsing, D.H.; Rosenbaum,
J. Biol. Chem. 270, 22522-22526, 1995
A;Title: Identification of a human type II receptor for bone morphogenetic protein-4 tha
A;Reference number: I55438; MUID:95403457; PMID:77673243
A;Accession: I55438
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/ENB
A;Molecule type: mRNA
A;Residues: 1-1038 <NOH>
A;Cross-references: NID:g907712; PIDN:BA09094.1; PID:g907713
R;Rosenzweig, B.L.; Imanura, T.; Okadome, T.; Cox, G.N.; Yamashita, H.; ten Dijke, P.; H
Proc. Natl. Acad. Sci. U.S.A. 92, 7632-7636, 1995

A;Title: Cloning and characterization of a human type II receptor for bone morphogenetic
A;Reference number: I37209; MUID:95372334; PMID:7644468
A;Accession: I37209
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-827, 'R', 829-1038 <ROS>
A;Cross-references: ENBL:Z48923; NID:g1009409; PIDN:CAA88759.1; PID:g1009410
C;Genetics:
A;Gene: GDB:BMPR2; BRK-3; T-ALK; BMPR3; BMPR-II
A;Cross-references: GDB:642243; OMIM:600799
A;Map position: 20pter-20qter
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;127-1038/Product: bone morphogenetic protein receptor II #status predicted <MAT>
F;127-150/Domain: extracellular #status predicted <EXT>
F;151-170/Domain: transmembrane #status predicted <TRM>
F;201-508/Domain: protein kinase homology <KIN>
F;209-217/Region: protein kinase ATP-binding motif
F;55,110,126/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 1.1%; Score 8; DB 2; Length 1038;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAQRS 336
Db 468 ENSLAQRS 475

RESULT 14
JC5527
bone morphogenetic protein type II receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 24-Sep-1999
C;Accession: JC5527
R;Beppu, H.; Minowa, O.; Miyazono, K.; Kawabata, M.
Biochem. Biophys. Res. Commun. 235, 499-504, 1997
A;Title: cDNA cloning and genomic organization of the mouse BMP type II receptor.
A;Reference number: JC5527; MUID:97350808; PMID:9207184
A;Accession: JC5527
A;Molecule type: mRNA
A;Residues: 1-1038 <BEP>
A;Cross-references: DDBJ:AF003942; NID:g2253704; PIDN:AAB63042.1; PID:g2253705
C;Comment: This protein is a serine/threonine kinase receptor that forms a heteromeric c
C;Genetics:
A;Gene: BMPR-II
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
F;1-26/Domain: signal sequence #status predicted <SIG>
F;151-174/Domain: transmembrane #status predicted <TRM>
F;201-508/Domain: protein kinase homology <KIN>
F;202-500/Domain: serine/threonine kinase #status predicted <STK>

Query Match 1.1%; Score 8; DB 2; Length 1038;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAQRS 336
Db 468 ENSLAQRS 475

RESULT 15
B71621
probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71621
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
P;Petter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71621
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2295 <GAR>
A:Cross-references: GB:AE001379; GB:AE001362; NID:g3845118; PIDN:AAC71827.1; PID:g384511
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0190c

Query Match 1.1; Score 8; DB 2; Length 2295;
Best Local Similarity 100.0%; Prad.No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 DSLRNSST 95
DB 1043 DSLRNSST 1050

Search completed: January 6, 2004, 09:53:41
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:48:30 ; Search time 17 seconds
(without alignments)
2099.604 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPFNRQWTSLSRVTA.....LSVEEQIKRNYDEDEEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	759	100.0	759	1 EPLI_HUMAN	Q9ub66 homo sapien
2	69	9.1	753	1 EPLI_MOUSE	Q9etg0 mus musculus
3	8	1.1	398	1 LH33_BRARE	Q90421 brachydanio
4	8	1.1	405	1 XSG3_CREEL	Q09379 caenorhabdi
5	8	1.1	415	1 MTDI_DESDN	P05302 desulfovibr
6	8	1.1	456	1 SELA_RHIME	P58226 rhizobium m
7	8	1.1	642	1 LIK2_CHICK	P53666 gallus gall
8	8	1.1	792	1 GYR2_AQUAE	Q67137 aquifex aco
9	8	1.1	1038	1 BMR2_HUMAN	Q13873 homo sapien
10	8	1.1	1038	1 BMR2_MOUSE	Q35607 mus musculus
11	7	0.9	85	1 YGFI_HAEIN	P44025 haemophilus
12	7	0.9	101	1 SU11_AERPE	Q9vbg9 aeropyrum p
13	7	0.9	149	1 HMGL_VICFA	P40620 vicia faba
14	7	0.9	156	1 PYR1_METAC	Q8thl3 methanosarc
15	7	0.9	156	1 PYR1_VETMA	Q8pxk6 methanosarc
16	7	0.9	159	1 CUP9_DROME	P27781 drosophila
17	7	0.9	178	1 RPOE_LISMO	Q8y494 listeria mo
18	7	0.9	195	1 GRPE_FRATU	P48204 francisella
19	7	0.9	198	1 GRPE_BACSH	Q69267 bacillus sp
20	7	0.9	198	1 VH23_SFVKA	Q9Q907 shope fibro
21	7	0.9	203	1 YP72_NAIZE	Q25737 zea mays (m
22	7	0.9	214	1 IPYR_ORISA	Q25337 oryza sativ
23	7	0.9	215	1 IPYR_HORVD	Q23979 hordeum vul
24	7	0.9	228	1 THIE_METAC	Q8tmd6 methanosarc
25	7	0.9	233	1 CYCX_NITEU	Q50926 nitrosomona
26	7	0.9	233	1 Y0D2_CABEL	P34594 caenorhabdi
27	7	0.9	249	1 PSE1_HUMAN	Q06323 h proteasom
28	7	0.9	249	1 PSE1_WACFA	P58238 macaca fasc
29	7	0.9	251	1 I433_FUCVE	Q39757 fucus vesic
30	7	0.9	251	1 A32B_HUMAN	Q92688 homo sapien
31	7	0.9	254	1 OAZ_DROME	P54361 drosophila
32	7	0.9	256	1 YP96_CABEL	Q09476 caenorhabdi
33	7	0.9	264	1 KSGA_FUSNN	Q8r6b1 f dimethyla

Q915u5 p dimethyla
P57854 pasteurella
P18150 streptomyce
P21135 schizosacch
Q15077 homo sapien
Q63371 rattus norv
P14594 pisum sativ
O02197 schistosoma
P37622 escherichia
Q92c17 listeria in
Q8y7q2 listeria mo
Q9a9e4 caulobacter

ALIGNMENTS

RESULT 1

EPLI_HUMAN
ID AC Q9UHB6; Q9BVF2; Q9H8J1; Q9HBN5; Q9NX96; Q9NXC3; Q9NXU6; Q9POH8;
AC Q9UHB5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Epithelial protein lost in neoplasm.
GN EPLIN OR SREBP3.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=20087188; PubMed=10618726;
RA Maul R.S., Chang D.D.;
RT "EPLIN, epithelial protein lost in neoplasm";
RL Oncogene 18:7838-7841(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RX MEDLINE=20267849; PubMed=10806352;
RA Chen S., Maul R.S., Kim H.R., Chang D.D.;
RT "Characterization of the human EPLIN (Epithelial protein lost in neoplasm) gene reveals distinct promoters for the two EPLIN isoforms";
RL Gene 248:69-76(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM BETA).
RT TISSUE=Uterus;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansoerge W., Boecher M., Bloecker H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N., Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs";
RL Genome Res. 11:422-435(2001).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA; BETA AND 3).
RC TISSUE=Colon, Hepatoma, and Placenta;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Hypothalamus;
RX MEDLINE=20402571; PubMed=10931946;
RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X., Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H., Gu B.-W., Dai N., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,

RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,
RT Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.,
RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal
RT axis and full-length cDNA cloning.",
RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548 (2000).
RN [6]
RN
RP SEQUENCE OF 232-759 FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.,
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.",
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RN
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE:Colon, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.W., Krzywinski M.I., Skalska U., Smalius D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. This cytoskeletal protein co-
CC localizes with actin stress fibers and focal adhesion plaques.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=Beta;
CC IsoId=Q9UHB6-1; Sequence=Displayed;
CC Name=Alpha;
CC IsoId=Q9UHB6-2; Sequence=VSP_003116;
CC Name=3;
CC IsoId=Q9UHB6-3; Sequence=VSP_003117;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta, kidney,
CC pancreas, prostate, ovary, spleen and heart. Also detected in
CC lung, liver, brain, skeletal muscle, thymus, testis and intestine.
CC Not detected in leukocytes. Epln-beta levels are generally very
CC low. Epln-alpha is abundant in epithelial cells from mammary
CC gland, prostate and in normal oral keratinocytes. Levels are low
CC in aortic endothelial cells and dermal fibroblasts, and not
CC detectable in myocardium.
CC -!- INDUCTION: Down-regulated in some cancer cell lines. Epln-alpha
CC is induced by serum. Epln-beta is constitutively expressed.
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -!- CAUTION: Ref.4 (BAA91120) sequence differs from that shown due to
CC a frameshift in position 697.
CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts
CC in positions 365 and 662.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC
DR EMBL; AF198454; AAF23755.1; -
DR EMBL; AF198455; AAF23756.1; -
DR EMBL; ALI36911; CAB66845.1; -
DR EMBL; AK000372; BAA91120.1; ALT_FRAME.
DR EMBL; AK000335; BAA91092.1; -
DR EMBL; AK023849; BAB14625.1; -
DR EMBL; AK000057; BAA90914.1; -
DR EMBL; AF157325; AAF67491.1; -
DR EMBL; AF218025; AAG17267.1; ALT_FRAME.
DR EMBL; BC001247; AAH01247.1; -
DR EMBL; BC010664; AAH10664.1; -
DR HSP; P32965; ICTL.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
KW Cytoskeleton; LIM domain; Metal-binding; Zinc; Alternative splicing.
FT DOMAIN 388 448 LIM.
FT VARSPLIC 1 160 Missing (in isoform Alpha).
FT VARSPLIC 1 302 Missing (in isoform 3).
FT CONFLICT 344 344 R -> PG (IN REF. 1; AAF23756).
FT CONFLICT 381 381 MISSING (IN REF. 4; BAA90914 AND 7).
FT CONFLICT 415 415 F -> L (IN REF. 6).
FT CONFLICT 463 463 D -> G (IN REF. 4; BAA90914).
FT CONFLICT 491 491 P -> Q (IN REF. 5).
FT CONFLICT 520 521 DK -> NR (IN REF. 5).
SQ SEQUENCE 759 AA; 85225 MW; 996378AFD3B003D5 CRC64;

Query Match 100.0%; Score 759; DB 1; Length 759;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESSPFNRQWTSLSLRVTAKLSLVNKKSAIIVEIFSKYKAAAEETNMKKRSNTNL 60
DB 1 MESSPFNRQWTSLSLRVTAKLSLVNKKSAIIVEIFSKYKAAAEETNMKKRSNTNL 60

QY 61 SQHFRKGTTLVLKKWENPGLCAESHTSLRNSSTIIRHRADHPAETVSHAASGAKADO 120
DB 61 SQHFRKGTTLVLKKWENPGLCAESHTSLRNSSTIIRHRADHPAETVSHAASGAKADO 120

QY 121 EQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHSSTESKMKMENCLESRHEVEKSEISEN 180
DB 121 EQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHSSTESKMKMENCLESRHEVEKSEISEN 180

QY 181 TDASGKIEKYNVPLNRLKXMPKGEPTQTTLRAOSRSASGRKISSENSYSLDDLEIGPGQ 240
DB 181 TDASGKIEKYNVPLNRLKXMPKGEPTQTTLRAOSRSASGRKISSENSYSLDDLEIGPGQ 240

QY 241 LSSSTFDSEKSNRRNLPLRLSETSIKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
DB 241 LSSSTFDSEKSNRRNLPLRLSETSIKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300

QY 301 HXWEQKENVPPGPEVCITHQGEKISANENSLAVRSTPAEDDSRDSQVSEVQQVPHKPP 360
DB 301 HXWEQKENVPPGPEVCITHQGEKISANENSLAVRSTPAEDDSRDSQVSEVQQVPHKPP 360

QY 361 LSPDRASSLSSESPPKAMKXFOAPARETCVECKTVYPMERLLANQQVFIHSCFRCSYC 420
DB 361 LSPDRASSLSSESPPKAMKXFOAPARETCVECKTVYPMERLLANQQVFIHSCFRCSYC 420

QY 421 NNKLSLGYASLHRIYCKPHNQLFKSGKGYDGFGRPHKDLWASKNVEEILERPQAQ 480
DB 421 NNKLSLGYASLHRIYCKPHNQLFKSGKGYDGFGRPHKDLWASKNVEEILERPQAQ 480

QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASQOEKEDKPAETKKLRIAWPPPTLGS 540
DB 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASQOEKEDKPAETKKLRIAWPPPTLGS 540

KW Metal-binding; Zinc; Transcription regulation; Activator.
FT DOMAIN 28 78 LIM 1.
FT DOMAIN 87 141 LIM 2.
FT DNA_BIND 154 213 HOMEOBOX.
SQ SEQUENCE 398 AA; 44226 MW; F0298F0D1A725C1F CRC64;

Query Match 1.1%; Score 8; DB 1; Length 398;
Best Local Similarity 100.0%; Pred.No.9.8; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 368 SSLSESSP 375
Db 272 SSLSESSP 279
|||||

RESULT 4
ID YS63 CAEEL STANDARD; PRT; 405 AA.
AC Q09179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK675.3 in chromosome II.
GN ZK675.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Sims M.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46812; CAA86845.1; -;
DR PIR; T27971; T27971.
DR WormPep; ZK675.3; CE01722.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 47835 MW; 0F589D874E9E0E61 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 405;
Best Local Similarity 100.0%; Pred.No.10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 SKENEEI 474
Db 254 SKENEEI 261
|||||

RESULT 5
MTDI DESDN STANDARD; PRT; 415 AA.
AC P05302;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase DdeI (EC 2.1.1.73) (Cytosine-specific
DE methyltransferase DdeI) (M.DdeI).
GN DDEIM OR DDEM.
OS Desulfovibrio desulfuricans (strain Norway 4).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;

OC Desulfomicrobiaceae; Desulfomicrobium.
OX NCBI_TaxID=52561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88040453; PubMed=2823226;
RA Snyder L.A., Slatko B., Moran L., O'Donnell K.H., Brooks J.E.;
RT "Nucleotide sequence of the DdeI restriction-modification system and
RT characterization of the methylase protein.";
RL Nucleic Acids Res. 15:8249-8266(1987)
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CTNAG, CAUSES SPECIFIC METHYLATION ON C-1 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE DDEI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC adenosyl-L-homocysteine + DNA 5-methylcytosine
CC -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00449; CAA68505.1; -;
DR HSSP; P20589; 1DCT
DR REBASE; 3357; M.DdeI.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PRO0105; C5METTRFRASE.
DR TIGRFAMs; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
KW Transferase; Methyltransferase; Restriction system.
FT ACT_SITE 76 76 BY SIMILARITY.
SQ SEQUENCE 415 AA; 47081 MW; 34DF5968E0414EED CRC64;

Query Match 1.1%; Score 8; DB 1; Length 415;
Best Local Similarity 100.0%; Pred.No.10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 TVASFSQS 599
Db 296 TVASFSQS 303
|||||

RESULT 6
SELA RHIME
ID SELA RHIME STANDARD; PRT; 466 AA.
AC P58226;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1) (Cysteiny-
DE tRNA(Sec) selenium transferase) (Selenocysteine synthase)
DE (Selenocysteinyl-tRNA(Sec) synthase).
GN SELA OR RA0005 OR SMA0011.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Beck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: Converts seryl-tRNA(Sec UCA) to selenocysteinyl-tRNA(Sec
CC UCA) during selenoprotein biosynthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: L-seryl-tRNA(Sec) + selenophosphate = L-
CC selenocysteinyl-tRNA(Sec) + H2O + phosphate.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SELA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007195; AAK64663.1; -
DR PIR; E95262; E95262.
DR HAMAP; MF 00423; - 1.
DR InterPro; IPR004534; Sela.
DR Pfam; PF03841; Sela; 1.
DR TIGRFAMs; TIGR00474; sela; 1.
DR Transferase; Pyridoxal phosphate; Selenium; Plasmid;
KW Complete proteome.
FT BINDING 292 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 466 AA; 49028 MW; B20B8852451C2EAD CRC64;

Query Match 1.1%; Score 8; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 539 GSSGSALE 546
Db 411 GSSGSALE 418
|||||

RESULT 7
LIK2_CHICK ID LIK2_CHICK STANDARD; PRT; 642 AA.
AC P53666;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LIM domain kinase 2 (EC 2.7.1.-) (LIK2).
GN LIK2 OR LIMK.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RS TISSUE=Lung.
RX MEDLINE=9515277; PubMed=7852284;
RA Ohashi K., Toshima J., Tajima K., Nakamura T., Mizuno K.;
RT "Molecular cloning of a chicken lung cDNA encoding a novel protein
RL J. Biochem. 116:636-642(1994).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, AND
CC FAINTLY IN THE KIDNEY, LIVER, BRAIN, SPLEEN, GIZZARD, AND
CC INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; D28310; BAA05372.1; -
DR PIR; JP0079; JP0079.
DR HSP; P32965; ICTL.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000094; LIM; 2.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00132; LIM; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR PROSITE; PS00023; LIM_DOMAIN_2; 2.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW LIM domain; Metal-binding; Zinc.
FT DOMAIN 12 63 LIM 1.
FT DOMAIN 72 124 LIM 2.
FT DOMAIN 152 239 PDZ.
FT DOMAIN 331 608 PROTEIN_KINASE.
FT NP_BIND 337 345 ATP (BY SIMILARITY).
FT BINDING 360 360 ATP (BY SIMILARITY).
FT ACT_SITE 451 451 BY SIMILARITY.
SQ SEQUENCE 642 AA; 72467 MW; 6D3E500B8C3A8277 CRC64;

```

Query Match 1.1%; Score 8; DB 1; Length 642;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 411 HISCRCFS 418
 Db 34 HISCRCFS 41
 |||||

```

R"RESULT 8
GARB_AQUAE ID GARB_AQUAE STANDARD; PRT; 792 AA.
AC O67137;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN GYRB OR AQ_1026.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]_TaxID=63363;
RP SEQUENCE FROM N.A.
RS STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujaay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

```

```

CC ENZYME FORMS AN A2B2 TETRAMER.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000719; AAC07098.1; -.
CC DR PIR; F70388; F70388.
CC DR HSSP; P06982; 1AJ6.
CC DR InterPro; IPR001594; ATPbind_ATPase.
CC DR InterPro; IPR002286; DNA_gyrase_C.
CC DR InterPro; IPR000565; DNA_gyrase.
CC DR InterPro; IPR001241; DNA_topoisomII.
CC DR InterPro; IPR006171; Toprim_dom.
CC DR Pfam; PF00204; DNA_gyrase; 1.
CC DR Pfam; PF00986; DNA_gyraseB_C; 1.
CC DR Pfam; PF02518; HATPase_C; 1.
CC DR Pfam; PF01751; Toprim; 1.
CC DR PRINTS; PR00418; TP12FAMILY.
CC DR ProDom; PD149633; DNA_gyraseB_C; 1.
CC DR SMART; SM00387; HATPase_C; 1.
CC DR SMART; SM00433; TOP2C; 1.
CC DR TIGRFS; TIGR01059; gyrB; 1.
CC DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
CC DR Topoisomerase; Isomerase; ATP-binding; Complete proteome.
CC KW SEQUENCE 792 AA; 90530 MW; 07DC02DF293FF9F3 CRC64;
CC SQ

```

```

Query Match 1.1%; Score 8; DB 1; Length 792;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 572 DLDLKKLR 579
Db 486 DLDLKKLR 493
|||||

```

```

RESULT 9
BMR2_HUMAN STANDARD; PRT; 1038 AA.
AC Q1873; Q18569;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bone morphogenetic protein receptor type II precursor (EC 2.7.1.37)
DE (BMP type II receptor) (BMPR-II).
GN BMPR2 OR PPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Substantia nigra;
RX MEDLINE=95372334; PubMed=7644468;
RA Rosenzweig B.L., Imamura T., Okadome T., Cox G.N., Yamashita H.,
RA ten Dijke P., Heldin C.H., Miyazono K.;
RT "Cloning and characterization of a human type II receptor for bone
RT morphogenetic proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 92:7632-7636(1995).
[2]_
RN SEQUENCE FROM N.A.
RP TISSUE=Skin fibroblast;
RX MEDLINE=9403457; PubMed=7673243;
RA Nohno T., Ishikawa T., Saito T., Hosokawa K., Noji S., Wosing D.H.,
RA Rosenbaum J.S.;
RT "Identification of a human type II receptor for bone morphogenetic
RT protein-4 that forms differential heteromeric complexes with bone
RT morphogenetic protein type I receptors."

```

```

J. Biol. Chem. 270:22522-22526(1995).
[3]_
RN SEQUENCE FROM N.A.
RX MEDLINE=95197572; PubMed=7890683;
RA Kawabata M., Chytil A., Moses H.L.;
RT "Cloning of a novel type II serine/threonine kinase receptor through
RT interaction with the type II transforming growth factor-beta
RT receptor."
RL J. Biol. Chem. 270:5625-5630(1995).
[4]_
RN VARIANTS PPH1 GLN-491 AND TRP-491.
RX MEDLINE=20395844; PubMed=10903931;
RA Deng Z., Morse J.H., Slager S.L., Quervo N., Moore K.J., Venetos G.,
RA Kalachikov S., Cayanis E., Fischer S.G., Barst R.J., Hodge S.E.,
RA Knowles J.A.;
RT "Familial primary pulmonary hypertension (gene PPH1) is caused by
RT mutations in the bone morphogenetic protein receptor-II gene."
RL Am. J. Hum. Genet. 67:737-744 (2000).
[5]_
RN VARIANTS PPH1 TYR-60; TYR-117 AND ARG-483.
RX MEDLINE=20473811; PubMed=11015450;
RA Thomson J.R., Machado R.D., Paucululo M.W., Morgan N.V., Humbert M.,
RA Elliott G.C., Ward K., Yacoub M., Mikhail G., Rogers P., Newman J.H.,
RA Wheeler L., Higenbottam T., Gibbs J.S.R., Egan J., Crozier A.,
RA Peacock A., Allcock R., Corris P., Loyd J.E., Trembath R.C.,
RA Nichols W.C.;
RT "Sporadic primary pulmonary hypertension is associated with germline
RT mutations of the gene encoding BMPR-II, a receptor member of the
RT TGF-beta family."
RL J. Med. Genet. 37:741-745(2000).
[6]_
RN VARIANTS PPH1 TRP-118; TYR-347 AND GLY-485.
RX MEDLINE=20428187; PubMed=10973254;
RA Lane K.B., Machado R.D., Paucululo M.W., Thomson J.R., Aldred M.,
RA Phillips J.A. III, Loyd J.E., Nichols W.C., Trembath R.C., Gaddipati R.,
RA Brannon C.A., Conneally P.M., Foroud T., Fretwell N., Gaddipati R.,
RA Koller D., Loyd E.J., Morgan N.V., Newman J.H., Prince M.A.,
RA Villano Gueell C., Wheeler L.;
RT "Heterozygous germline mutations in BMPR2, encoding a TGF-beta
RT receptor, cause familial primary pulmonary hypertension."
RL Nat. Genet. 26:81-84(2000).
[7]_
RN VARIANTS PPH1 ARG-123; SER-123; ARG-420 AND THR-512, VARIANT ASP-224,
AND CHARACTERIZATION OF VARIANT PPH1 GLY-485.
RX MEDLINE=21063176; PubMed=1115378;
RA Machado R.D., Paucululo M.W., Thomson J.R., Lane K.B., Morgan N.V.,
RA Wheeler L., Phillips J.A. III, Newman J.H., Williams D., Galie N.,
RA Manes A., McNeil K., Yacoub M., Mikhail G., Rogers P., Corris P.,
RA Humbert M., Donnai D., Martensson G., Tranebjærg L., Loyd J.E.,
RA Trembath R.C., Nichols W.C.;
RT "BMPR2 haploinsufficiency as the inherited molecular mechanism for
RT primary pulmonary hypertension."
RL Am. J. Hum. Genet. 68:92-102(2001).
CC -!- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4.
CC BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS
CC FOR BMPs.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTORS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART AND LIVER.
CC -!- DISEASE: Defects in BMPR2 are the cause of primary pulmonary
CC hypertension (PPH1) [MIM:178600]; a rare autosomal dominant
CC disorder characterized by pleiomorphic lesions of proliferating
CC endothelial cells in pulmonary arterioles. The lesions lead to
CC elevated pulmonary arterial pressure, right ventricular failure,
CC and death. The disease can occur from infancy throughout life and
CC it has a mean age at onset of 36 years. Penetrance is reduced.
CC Although familial PPH1 is rare, cases secondary to known
CC etiologies are more common and include those associated with the
CC appetite-suppressant drugs.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGF-beta RECEPTOR SUBFAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
EMBL; Z48923; CAA88759.1; -
EMBL; D50516; BAA09094.1; -
EMBL; U20165; AAC50105.1; -
PIR; I38935; I38935..
Genew; HGNC:1078; BMPR2..
MIN; 600799; -
MIN; 178600; -
GO; GO:0005887; C: integral to plasma membrane; TAS.
GO; GO:0005515; F: protein binding activity; TAS.
GO; GO:0007178; P: transmembrane receptor protein serine/threonine/threo. .; TAS.
InterPro; IPR000472; Activin_rec.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002290; Ser_thr_pkinase.
Pfam; PF01064; Activin_rec; 1.
Pfam; PF00069; Pkinase; 1.
ProDom; PD000001; Prot_kinase; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
Transmembrane; Signal; Polymorphism; Disease mutation.
FT SIGNAL 1 26
FT CHAIN 27 1038
FT DOMAIN 27 150 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 151 171 POTENTIAL.
FT DOMAIN 172 1038 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 203 504 PROTEIN_KINASE.
FT NP_BIND 209 217 ATP (BY SIMILARITY).
FT BINDING 230 230 ATP (BY SIMILARITY).
FT ACT_SITE 333 333 BY SIMILARITY.
FT DOMAIN 547 550 POLY-SER.
FT DOMAIN 610 618 POLY-THR.
FT DOMAIN 901 908 POLY-ASN.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 60 60 C -> Y (in PPH1).
FT /FTID-VAR_013670.
FT VARIANT 117 117 /FTID-VAR_013671.
FT VARIANT 118 118 C -> W (in PPH1).
FT /FTID-VAR_013672.
FT VARIANT 123 123 C -> R (in PPH1).
FT /FTID-VAR_013673.
FT VARIANT 123 123 C -> S (in PPH1).
FT /FTID-VAR_013674.
FT VARIANT 224 224 E -> D.
FT /FTID-VAR_013675.
FT VARIANT 347 347 C -> Y (in PPH1).
FT /FTID-VAR_013676.
FT VARIANT 420 420 C -> R (in PPH1).
FT /FTID-VAR_013677.
FT VARIANT 483 483 C -> R (in PPH1; sporadic).
FT /FTID-VAR_013678.
FT VARIANT 485 D -> G (in PPH1; complete loss of function).
FT /FTID-VAR_013679.
FT VARIANT 491 R -> Q (in PPH1; sporadic).
FT /FTID-VAR_013680.
FT VARIANT 491 R -> W (in PPH1).
FT /FTID-VAR_013681.
FT VARIANT 512 K -> T (in PPH1).
FT /FTID-VAR_013682.
FT VARIANT 519 N -> K (in PPH1).
```

```
FT CONFLICT 828 828 /FTID-VAR_013683.
FT G -> R (in REF. 1).
SQ SEQUENCE 1038 AA; 115201 MW; 1389923CE574B913 CRC64;
Query Match 1.1%; Score 8; DB 1; Length 1038;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 329 ENSLAVERS 336
Db 468 ENSLAVERS 475
RESULT 10
BMR2_MOUSE
ID BMR2_MOUSE STANDARD; PRT; 1038 AA.
AC Q35607;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bone morphogenetic protein receptor type II precursor (EC 2.7.1.37)
DE (BMP type II receptor) (BMPR-II) (BRK-3).
DE BMPR2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97350808; PubMed=9207184;
RA Beppu H., Minowa O., Miyazono K., Kawabata M.;
RT "cDNA cloning and genomic organization of the mouse BMP type II
RT receptor."
RL Biochem. Biophys. Res. Commun. 235:499-504 (1997).
RL [2]
RP SEQUENCE FROM N.A.
RA Whitaker G.B., Koenig B.B., Ting J., Tiesman J.P., Limberg A.L.,
RA Grant R.A., Begley K.B., Rosenbaum J.S.;
RT "Identification of BMP receptor complexes with differential signaling
RT properties and ligand binding profiles."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4.
CC BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS
CC FOR BMPs.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTORS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGF-beta RECEPTOR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL; AF003942; AAB63042.1; -
EMBL; U78048; AAB87638.1; -
PIR; JC5527; JC5527.
MGD; MGI:1095407; Empr2.
GO; GO:0009928; C: cell surface (sensu Magnoliophyta); IDA.
InterPro; IPR000472; Activin_rec.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002290; Ser_thr_pkinase.
Pfam; PF01064; Activin_rec; 1.
ProDom; PD000001; Prot_kinase; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
```



```

KW Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1038
FT
FT
FT DOMAIN 27 150
FT TRANSMEM 151 171
FT DOMAIN 172 1038
FT DOMAIN 172 1038
FT DOMAIN 203 504
FT NP BIND 209 217
FT BINDING 230 230
FT ACT_SITE 333 333
FT DOMAIN 191 194
FT DOMAIN 547 550
FT DOMAIN 610 618
FT DOMAIN 901 908
FT CARBOHYD 55 55
FT CARBOHYD 110 110
FT CARBOHYD 126 126
FT CARBOHYD 126 126
SQ SEQUENCE 1038 AA; 115019 MW; 4106945DC63250E1 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1038;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAVERS 336
DB 468 ENSLAVERS 475
|||||

RESULT 11
YGFY_HAEIN STANDARD; PRT; 85 AA.
AC P44025.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0627.
GN HI0627.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu D.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512 (1995).
CC -!- SIMILARITY: STRONG, TO E.COLI YGFY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32745; AAC22287.1; -
CC PIR; G64010; G64010.
CC TIGR; HI0627; -
CC InterPro; IPR005631; DUF339.
CC Pfam; PF03937; TPR_div1; 1.

```

```

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 85 AA; 10430 MW; AAE707492162A1C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KSQDEL 731
DB 63 KSQDEL 69
|||||

RESULT 12
SUI1_AERPE STANDARD; PRT; 101 AA.
AC Q5V6G9.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein translation factor SUI1 homolog.
GN APE1629.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takehashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Koshida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101 (1999).
CC -!- SIMILARITY: BELONGS TO THE SUI1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000062; BAA80629.1; -
CC PIR; H72542; H72542.
CC HSSP; F08245; 1DIR.
CC HAMAP; MF_00604; -; 1.
CC InterPro; IPR005872; SUI1_arch_bact.
CC InterPro; IPR001950; TIF_SUI1.
CC Pfam; PF01253; SUI1; 1.
CC TIGRFAMs; TIGR01158; SUI1_rel; 1.
CC PROSITE; PS01118; SUI1_1; 1.
CC PROSITE; PS0296; SUI1_2; 1.
KW Protein biosynthesis; Translation regulation; Complete proteome.
SQ SEQUENCE 101 AA; 11470 MW; 80F377A4B658D636 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 740 LSVEEQI 746
DB 18 LSVEEQI 24
|||||

RESULT 13
HMGL_VICFA
ID "HMGL_VICFA STANDARD; PRT; 149 AA.

```

P40620;
 01-FEB-1995 (Rel. 31, Created)
 01-FEB-1995 (Rel. 31, Last sequence update)
 15-SEP-2003 (Rel. 42, Last annotation update)
 HMGI/2-like protein.
 Vicia faba (Broad bean).
 Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
 NCBI_TaxID=3906;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=cv. Minor; TISSUE=Cotyledon;
 MEDLINE=9403341; PubMed=8219095;
 Grasser K., Wohlfarth T., Baumstein H., Felix G.;
 "Comparative analysis of chromosomal HMG proteins from monocotyledons
 and dicotyledons";
 Plant Mol. Biol. 23:619-625(1993).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: BELONGS TO THE HMGI/HMG2 PROTEIN FAMILY.
 CC -!- SIMILARITY: Contains 1 HMG box domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; Z21703; CAB37859.1; -;
 PIR; S39556; S39556.
 DR HSSP; P07155; IHMF.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF0505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS0118; HMG_BOX_2; 1.
 DR Nuclear protein; DNA-Binding.
 FT DNA_BIND 45 114 HMG_BOX.
 FT DOMAIN 121 149 ASP/GLU-RICH (ACIDIC).
 FT SEQUENCE 149 AA; 16659 MW; FE52DF5E68E0B3D CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 753 DEDEEE 759
 DB 135 DEDEEE 141
 [1]
 RESULT 14
 ID PYRI METAC STANDARD; PRT; 156 AA.
 AC Q8THL3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aspartate carbamoyltransferase regulatory chain.
 GN PYRI OR MA4501.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 FitchHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 Linton L., McEwan P., McKernan K., Talanas J., Tirrell A., Ye W.,
 Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,

Hedderich R., Ingram-Smith C., Kuettner H.C., Krzyzski J.A.,
 Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 Metcalf W.W., Birren B.;
 "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity";
 Genome Res. 12:532-542(2002).
 CC -!- FUNCTION: Involved in allosteric regulation of aspartate
 carbamoyltransferase (By similarity).
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBUNIT: Contains catalytic and regulatory chains.
 CC -!- SIMILARITY: BELONGS TO THE PYRI FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; AE011170; AM07841.1; -;
 HAMAp; MF_00002; -; 1.
 DR InterPro; IPR002801; Pyri.
 DR Pfam; PF01948; Pyri; 1.
 DR Pfam; PF02748; Pyri_C; 1.
 DR ProDom; PD006194; Pyri; 1.
 DR TIGRFAMs; TIGR00240; Atcase_reg; 1.
 KW Pyrimidine biosynthesis; Zinc; Complete proteome.
 FT METAL 109 109 ZINC (BY SIMILARITY).
 FT METAL 114 114 ZINC (BY SIMILARITY).
 FT METAL 140 140 ZINC (BY SIMILARITY).
 FT METAL 143 143 ZINC (BY SIMILARITY).
 FT SEQUENCE 156 AA; 17336 MW; 51DF22654D3B83A1 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 156;
 Best Local Similarity 100.0%; Pred No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 738 KELSVEE 744
 DB 64 KELSVEE 70
 [1]
 RESULT 15
 ID PYRI METWA STANDARD; PRT; 156 AA.
 AC Q8PXK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aspartate carbamoyltransferase regulatory chain.
 GN PYRI OR MW1212.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 Martinez-Arias R., Henne A., Wierzer A., Baumer S., Jacobi C.,
 Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
 Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 Fritz H.-J., Gottschalk G.;
 "The genome of Methanosarcina mazei: evidence for lateral gene
 transfer between Bacteria and Archaea";
 J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -!- FUNCTION: Involved in allosteric regulation of aspartate
 carbamoyltransferase (By similarity).

```
CC -!- SUBUNIT: Contains catalytic and regulatory chains.
CC -!- SIMILARITY: BELONGS TO THE PYRI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013351; AAM30908.1; -
DR HAMAP; MF 00002; -; 1.
DR InterPro; IPR002801; Pyri.
DR Pfam; PF01948; Pyri.1.
DR Pfam; PF02748; Pyri.C; 1.
DR PRODOM; PD006194; Pyri; 1.
DR TIGRFAMs; TIGR00240; AtCase reg; 1.
KW Pyrimidine biosynthesis; Zinc; Complete proteome.
FT METAL 109 109 ZINC (BY SIMILARITY).
FT METAL 114 114 ZINC (BY SIMILARITY).
FT METAL 140 140 ZINC (BY SIMILARITY).
FT METAL 143 143 ZINC (BY SIMILARITY).
SQ SEQUENCE 156 AA; 17011 MW; B77F6C12F578A962 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 156;
Best Local Similarity 100.0%; Pred.No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 738 KELSVEE 744
Db 64 KELSVEE 70
|||||
```

Search completed: January 6, 2004, 09:52:13
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:49:05 ; Search time 41 seconds
(without alignments)
4777.120 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSFPNRRQWTSLSLRVTA.....LSVEEQIKRNYDEDEDEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	9.1	593	11 Q8CD09	Q8cd09 mus musculus
2	69	9.1	593	11 Q8BT15	Q8bt15 mus musculus
3	69	9.1	753	11 Q8X2H0	Q8x2h0 mus musculus
4	69	9.1	753	11 Q8C7S2	Q8c7s2 mus musculus
5	65	8.6	593	11 Q8C3R7	Q8c3r7 mus musculus
6	15	2.0	519	4 Q8N7Z0	Q8n7z0 homo sapien
7	15	2.0	951	6 Q8HXJ9	Q8hxj9 macaca fasc
8	14	1.8	127	4 Q8BT23	Q8bt23 homo sapien
9	14	1.8	127	4 Q8GS91	Q8gs91 homo sapien
10	14	1.8	128	11 Q8BGB5	Q8bgb5 mus musculus
11	14	1.8	629	13 Q8DEY8	Q8dey8 brachydanio
12	9	1.2	647	12 Q8DWB9	Q8dwb9 rat cytoneme
13	8	1.1	101	11 Q62493	Q62493 mus musculus
14	8	1.1	120	2 Q9ALE6	Q9ale6 uncultured
15	8	1.1	122	17 Q97Z14	Q97z14 sulfolobus
16	8	1.1	135	6 Q8HXR3	Q8hxr3 bos taurus

17	8	1.1	148	9 Q9B095	Q9b095 mycobacteri
18	8	1.1	169	17 Q8TV97	Q8tv97 methanopyru
19	8	1.1	231	10 Q8O964	Q8o964 arabidopsis
20	8	1.1	234	17 Q9YG41	Q9yg41 aeropyrum p
21	8	1.1	279	5 Q9V5Z4	Q9v5z4 drosophila
22	8	1.1	297	4 Q9A6A1	Q9a6a1 homo sapien
23	8	1.1	354	5 Q76Z76	Q76z76 drosophila
24	8	1.1	358	4 Q96BE7	Q96be7 homo sapien
25	8	1.1	365	11 Q8BHK0	Q8bhk0 mus musculu
26	8	1.1	374	13 Q42N86	Q42n86 brachydanio
27	8	1.1	386	10 Q9FN10	Q9fn10 arabidopsis
28	8	1.1	388	17 Q9XEY0	Q9xeY0 aeropyrum p
29	8	1.1	403	10 Q42381	Q42381 arabidopsis
30	8	1.1	413	10 Q64815	Q64815 arabidopsis
31	8	1.1	445	16 Q99YL0	Q99yl0 streptococc
32	8	1.1	445	16 Q8NZY9	Q8nzy9 streptococc
33	8	1.1	445	16 Q8K6D5	Q8k6d5 streptococc
34	8	1.1	459	5 Q9VCB8	Q9vcb8 drosophila
35	8	1.1	459	5 Q9LV69	Q9lv69 arabidopsis
36	8	1.1	473	16 Q8DLI3	Q8dli3 synechococc
37	8	1.1	530	4 Q13161	Q13161 homo sapien
38	8	1.1	682	10 Q94GH0	Q94gh0 oryza sativ
39	8	1.1	730	5 Q9VC06	Q9vc06 drosophila
40	8	1.1	784	16 Q9HXB2	Q9hxb2 pseudomonas
41	8	1.1	829	10 Q8LIY9	Q8liy9 oryza sativ
42	8	1.1	874	5 Q95X65	Q95x65 caenorhabdi
43	8	1.1	942	5 Q95X66	Q95x66 caenorhabdi
44	8	1.1	1026	10 Q9SS79	Q9ss79 arabidopsis
45	8	1.1	1031	13 Q42124	Q42124 gallus gall

ALIGNMENTS

RESULT 1

Q8CD09 PRELIMINARY; PRT; 593 AA.
AC Q8CD09;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12468851;
RA The FANTOM Consortium,
RA the RIXEN Genome Exploration Research Group Phase I & II Team;
RT 'Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.'
RL Nature 420:563-573(2002).
DR EMBL; AK031698; BAC27520.1;
SQ SEQUENCE 593 AA; 66017 MW; 122AB8AA07E1FC0A CRC64;

Query Match 9.1%; Score 69; DB 11; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYMERLLANQQVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHNQLPKSK 449

DB 228 CVEQKTVYMERLLANQQVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHNQLPKSK 287

QY 450 GNYDEGEFGH 458

DB 288 GNYDEGEFGH 296

RESULT 2

Q8BT15

```
ID Q8BT15 PRELIMINARY; PRT; 593 AA.
AC Q8BT15;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK028186; BAC25798.1; -.
SQ SEQUENCE 593 AA; 66026 MW; 3F2AB8AA07F7BE6E CRC64;

Query Match 9.1%; Score 69; DB 11; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 449
DB 228 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 287
QY 450 GNYDEGFGH 458
DB 288 GNYDEGFGH 296
QY 450 GNYDEGFGH 458
DB 288 GNYDEGFGH 296

RESULT 3
Q8K2H0 PRELIMINARY; PRT; 753 AA.
ID Q8K2H0;
AC Q8K2H0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; BC031490; AAH31490.1; -.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SMO0132; LIM; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
DR PROSITE; PS0023; LIM DOMAIN 2; 1.
DR LIM domain; Metal-Binding; Zinc.
SQ SEQUENCE 753 AA; 84059 MW; 79F98E47C100CF22 CRC64;

Query Match 9.1%; Score 69; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.7e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 449
DB 388 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 447
QY 450 GNYDEGFGH 458
DB 288 GNYDEGFGH 296
```

```
Db 448 GNYDEGFGH 456

RESULT 4
Q8C7S2 PRELIMINARY; PRT; 753 AA.
ID Q8C7S2;
AC Q8C7S2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049350; BAC33699.1; -.
SQ SEQUENCE 753 AA; 84089 MW; 39571A17DF21F2C0 CRC64;

Query Match 9.1%; Score 69; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.7e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 449
DB 388 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 447
QY 450 GNYDEGFGH 458
DB 448 GNYDEGFGH 456

RESULT 5
Q8C3R7 PRELIMINARY; PRT; 593 AA.
ID Q8C3R7;
AC Q8C3R7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK085065; BAC39353.1; -.
SQ SEQUENCE 593 AA; 66057 MW; FFB3E6B9506A61A CRC64;

Query Match 8.6%; Score 65; DB 11; Length 593;
Best Local Similarity 100.0%; Pred. No. 2.3e-59;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 449
DB 228 CVEQKTVYPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHFNQLFKSK 287
QY 450 GNYDE 454
```

```
Db 288 GNYDE 292
|||||
RESULT 6
Q8N7Z0 PRELIMINARY; PRT; 519 AA.
AC Q8N7Z0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-4AR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ40200.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Wagaatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL: AK097519; BAC05086.1; -.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 1.
DR ProDom: PD000094; LIM; 1.
DR SMART: SM00132; LIM; 1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 519 AA; 58797 MW; 69DE84BBE60E9106 CRC64;

Query Match 2.0%; Score 15; DB 4; Length 519;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFH 458
|||||
Db 456 QLFKSKGNYDEGFH 470

RESULT 7
Q8HXJ9 PRELIMINARY; PRT; 951 AA.
AC Q8HXJ9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Tetso K., Suzuki Y., Sugano S., Hashimoto K.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human

Chromosomes."
RL Gene 275:31-37(2001).
DR EMBL: AB097518; BAC41743.1; -.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 107682 MW; 42A3F8DEC05274B2 CRC64;

Query Match 2.0%; Score 15; DB 6; Length 951;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFH 458
|||||
Db 469 QLFKSKGNYDEGFH 493

RESULT 8
Q8BT23 PRELIMINARY; PRT; 127 AA.
AC Q8BT23;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ34982.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Wagaatsuma M.,
RA Wagaatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL: BC004400; AAH04400.1; -.
DR EMBL: AK092301; BAC03855.1; -.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 1.
DR ProDom: PD000094; LIM; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14070 MW; B63174FCF0486956 CRC64;

Query Match 1.8%; Score 14; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFH 457
|||||
Db 94 QLFKSKGNYDEGFH 107

RESULT 9
Q96S91 PRELIMINARY; PRT; 127 AA.
AC Q96S91;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Zhang W., Zhang M., Wan T., Cao X.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL; AY037154; AAK67634.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14142 MW; B63014FEF0486954 CRC64;

Query Match 1.8%; Score 14; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFSKSGNYDEGFG 457
Db 94 QLFSKSGNYDEGFG 107
|||||

RESULT 10
Q8BG5 PRELIMINARY; PRT; 128 AA.
AC Q8GB5;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Hypothetical LIM domain/LIM domain profile/cytochrome c family
DE heme-binding site containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Body, Hippocampus, and Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK012581; BAC25371.1; -.
DR EMBL; AK032430; BAC27866.1; -.
DR EMBL; AK049809; BAC33928.1; -.
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 14237 MW; ADF9161771331D13 CRC64;

Query Match 1.8%; Score 14; DB 11; Length 128;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFSKSGNYDEGFG 457
Db 95 QLFSKSGNYDEGFG 108
|||||

RESULT 11
Q9DEY8

ID Q9DEY8 PRELIMINARY; PRT; 629 AA.
AC Q9DEY8;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE Hypothetical protein.
DE Cytokeleton-associated LIM domain protein.
GN EPLIN.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100452; PubMed=11179579;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
RT and comparison of mammalian and zebrafish EPLIN."
RL Gene 262:155-160(2001).
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL; AF307846; AAG31149.1; -.
DR HSP; P32965; ICTD.
DR ZFIN; ZDB-GENE-001120-1; eplin.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 629 AA; 70037 MW; C42341B024818C03 CRC64;

Query Match 1.8%; Score 14; DB 13; Length 629;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 KGYDEGFGHRPHK 462
Db 345 KGYDEGFGHRPHK 358
|||||

RESULT 12
Q9DWB9 PRELIMINARY; PRT; 647 AA.
ID Q9DWB9;
AC Q9DWB9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE Hypothetical protein.
DE Cytokeleton-associated LIM domain protein.
GN EPLIN.
OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=79700;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Maastricht;
RX MEDLINE=20366325; PubMed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome."
RL J. Virol. 74:7656-7665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript."
RL Virus Res. 69:119-130(2000).
DR EMBL; AF232689; AAF99171.1; -.
SQ SEQUENCE 647 AA; 71565 MW; 8C18506591522F6C CRC64;

Query Match 1.2%, Score 9; DB 12; Length 647;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 YDEDEDEE 759
DQ 447 YDEDEDEE 455

RESULT 13
Q62493 PRELIMINARY; PRT; 101 AA.
ID Q62493
AC Q62493
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Unknown protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fizes C., Levi-Acobas F., Depetris D., Mattei M.G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
DR EMBL; 278149; CAB01551.1; -.
FT NON_TER 1 1
FT NON_TER 101 101
SQ SEQUENCE 101 AA; 11373 MW; 32A53AEC6AA7C477 CRC64;

Query Match 1.1%; Score 8; DB 11; Length 101;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 LSPDSRAS 368
DQ 81 LSPDSRAS 88

RESULT 14
Q9ALE6 PRELIMINARY; PRT; 120 AA.
ID Q9ALE6
AC Q9ALE6
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Dissimilatory sulfite reductase subunit A (fragment).
GN DSRA.
OS uncultured sulfate-reducing bacterium UMTASdr828-28.
OC Bacteria; Proteobacteria; Deltaproteobacteria; environmental samples.
OX NCBI_TaxID=151117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21318708; PubMed=11425735;
RA Chang Y.J., Peacock A.D., Long P.E., Stephen J.R., McKinley J.P.,
RA Macnaughton S.J., Hussain A.K., Saxton A.M., White D.C.;
RT "Diversity and characterization of sulfate-reducing bacteria in
RT groundwater at a uranium mill tailings site.";
RL Appl. Environ. Microbiol. 67:3149-3160(2001).
DR EMBL; AY015533; AAG61216.1; -.
DR InterPro; IPR006067; NIK_Sir_4Fe4S.
DR Pfam; PF01077; NIK_Sir; 1.
FT NON_TER 1 1
FT NON_TER 120 120
SQ SEQUENCE 120 AA; 13189 MW; 328A679BDABFA964 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 120;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 LQSSGSAL 545
DQ 84 LQSSGSAL 91

RESULT 15
Q97214 PRELIMINARY; PRT; 122 AA.
ID Q97214
AC Q97214
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Hypothetical protein SS00922.
GN SS00922.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tostrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006712; AAK41204.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 122 AA; 14725 MW; 944101A92BD87F15 CRC64;

Query Match 1.1%; Score 8; DB 17; Length 122;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 IEKYNVPL 194
DQ 37 IEKYNVPL 44

Search completed: January 6, 2004, 09:53:07
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:49:46 ; Search time 13111 Seconds
(without alignments)
11560.534 Million cell updates/sec

Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcgcaggagcagtaggtg.....atttgaaaaaaaaaaaaa 3705

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.:

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3660.4	98.8	3664	6	AX086386	AX086386 Sequence
2	3660.4	98.8	3664	9	HSN800164	AL136911 Homo sapi
3	3643.8	98.3	3655	9	AF198454	AF198454 Homo sapi
4	3374.4	91.1	3550	9	AF198455	AF198455 Homo sapi
5	3262.2	88.0	3277	9	AK000335	AK000335 Homo sapi
6	2698.2	72.8	2783	6	AX017487	AX017487 Sequence
7	2698.2	72.8	2783	6	BD135177	BD135177 Human nuc
8	2680	72.3	3348	9	AF157325	AF157325 Homo sapi
9	2604.4	70.3	2667	9	AK000057	AK000057 Homo sapi
10	2515.4	67.9	2536	9	AK000372	AK000372 Homo sapi
11	2284.4	61.7	115345	9	AC008147	AC008147 Homo sapi
12	2141	57.8	2164	9	BC001247	BC001247 Homo sapi
13	2115.2	57.1	2121	9	BC010664	BC010664 Homo sapi
14	1721.4	46.5	1754	9	AF218025	AF218025 Homo sapi
15	1568.4	42.3	2207	6	BD160117	BD160117 Primer fo
16	1568.4	42.3	2207	9	AK023649	AK023649 Homo sapi
17	1560.8	42.1	2793	9	AK096172	AK096172 Homo sapi
18	1559	42.1	3232	10	BC031490	BC031490 Mus muscu
19	1514.6	40.9	2370	10	AF307845	AF307845 Mus muscu
20	1409.6	38.0	3997	10	AF307844	AF307844 Mus muscu
21	1158.4	31.3	28186	9	AC139016	AC139016 Homo sapi
22	991	26.7	1021	6	AX012180	AX012180 Sequence
23	806.4	21.8	4425	10	CGU22818	U22818 Cricetus
24	766.4	20.7	777	11	G56637	G56637 SHGC-102015
25	669.6	18.1	196301	2	AC134548	AC134548 Mus muscu
26	637.4	17.2	641	6	AX011639	AX011639 Sequence
27	605.4	16.3	237661	2	AC114454	AC114454 Rattus no
28	416	11.2	532	6	BD154971	BD154971 Primer fo
29	390	10.5	390	6	BD113230	BD113230 EST and e
30	377.6	10.2	389	6	B0025824	B0025824 Sequence
31	352.6	9.5	413	6	AR270117	AR270117 Sequence
32	343	9.3	343	6	AX378979	AX378979 Sequence
33	343	9.3	343	6	AX379294	AX379294 Sequence
34	310.2	8.4	195516	2	AC135476	AC135476 Rattus no
35	277.4	7.5	279	6	AR269930	AR269930 Sequence
36	240.8	6.5	250	11	G11512	G11512 human STS S
37	224	6.0	238	6	AR269818	AR269818 Sequence
38	212	5.7	247	6	AX184947	AX184947 Sequence
39	210.8	5.7	231	6	AX187815	AX187815 Sequence
40	210.8	5.7	255	6	AX186750	AX186750 Sequence
41	209.4	5.7	265	6	AX098162	AX098162 Sequence
42	199.2	5.4	296	6	B0027230	B0027230 Sequence
43	175.2	4.7	178	6	AX344839	AX344839 Sequence
44	147.4	4.0	2264	9	AK097519	AK097519 Homo sapi
45	145.8	3.9	3287	9	AB097518	AB097518 Macaca fa

ALIGNMENTS

RESULT 1
AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Wiemann,S.
TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 338 22-FEB-2001;
Pred. No. is the number of results predicted by chance to have a

FEATURES		German Human Genome Project (DE)	
source	Location/Qualifiers		
	1..3664		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	1210 a 734 c 809 g 911 t		
ORIGIN			
	Query Match	98.8%;	Score 3660.4; DB 6; Length 3664;
	Best Local Similarity	100.0%;	Pred. No. 0;
	Matches 3661; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
QY	44	GGCTAGGTAGACGCCGGGACCTGTGACAGGCTGTGACAGCGGACGAGAGAAAGCGGC	103
Db	1	GGCTAGGTAGACGCCGGGACCTGTGACAGGCTGTGACAGCGGACGAGAGAAAGCGGC	60
QY	104	TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTAAATAGACGG	163
Db	61	TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTAAATAGACGG	120
QY	164	CAATGGACCTCACTATCATTTAGGGGTAACAGCCAAAGAACTTTCTCTTGTCAACAAGAAC	223
Db	121	CAATGGACCTCACTATCATTTAGGGGTAACAGCCAAAGAACTTTCTCTTGTCAACAAGAAC	180
QY	224	AAGTCATCGGCTATTGTGAAATATTTCTCAAGTACCAAGAACGAGCTGAAGAAAACAAAC	283
Db	181	AAGTCATCGGCTATTGTGAAATATTTCTCAAGTACCAAGAACGAGCTGAAGAAAACAAAC	240
QY	284	ATGGAGAAGAGAGAGTAAACACCGAAATCTCTCCAGCACTTTAGAAAGGGGACCTCG	343
Db	241	ATGGAGAAGAGAGAGTAAACACCGAAATCTCTCCAGCACTTTAGAAAGGGGACCTCG	300
QY	344	ACTGTGTTAAAGAGAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT	403
Db	301	ACTGTGTTAAAGAGAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT	360
QY	404	CTACGGAACAGCAGCACTAGATTTAGGACAGAGCAGCACTCTCTCTGCTGAAGTGACA	463
Db	361	CTACGGAACAGCAGCACTAGATTTAGGACAGAGCAGCACTCTCTCTGCTGAAGTGACA	420
QY	464	AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGACTAGA	523
Db	421	AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGACTAGA	480
QY	524	CTCAGGTCACTCTGAAAGCCCTCTGTTTCAGGTCGATATCCCCACATCAAGGACGGTGAG	583
Db	481	CTCAGGTCACTCTGAAAGCCCTCTGTTTCAGGTCGATATCCCCACATCAAGGACGGTGAG	540
QY	584	GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATTGTCTAGGAGAAATCCAGG	643
Db	541	GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATTGTCTAGGAGAAATCCAGG	600
QY	644	CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAA	703
Db	601	CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAA	660
QY	704	TATAATGTTCCGCTGAACAGGCTTAAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT	763
Db	661	TATAATGTTCCGCTGAACAGGCTTAAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT	720
QY	764	AAGATTCTCCGGGCCAAAGCCGAAGTGCAGAGTGAAGAGAGATCTCTGAAAAACAGCTAT	823
Db	721	AAGATTCTCCGGGCCAAAGCCGAAGTGCAGAGTGAAGAGAGATCTCTGAAAAACAGCTAT	780
QY	824	TCTCTAGATGACCTGGAAATAGGCCAGGTGAGTTGTCATCTTACATTTGACTTCGAG	883
Db	781	TCTCTAGATGACCTGGAAATAGGCCAGGTGAGTTGTCATCTTACATTTGACTTCGAG	840
QY	884	AAAAATGAGAGTAGACGAAATCTGGAATCTCCAGCCCTCTCAGAAACCTCTATAAAGAT	943
Db	841	AAAAATGAGAGTAGACGAAATCTGGAATCTCCAGCCCTCTCAGAAACCTCTATAAAGAT	900

QY	944	CGAATGCCAAGTACCAAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT	1003
Db	901	CGAATGCCAAGTACCAAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT	960
QY	1004	GAGCTGAAGCCAGTGTGGGAAATCAAAATTCATAAAATGAGCAAAAGGAGATGTG	1063
Db	961	GAGCTGAAGCCAGTGTGGGAAATCAAAATTCATAAAATGAGCAAAAGGAGATGTG	1020
QY	1064	CCCCAGGTCTGTAGGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG	1123
Db	1021	CCCCAGGTCTGTAGGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG	1080
QY	1124	AATAGCCTGCGAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTGACTCCCAAGTTAAG	1183
Db	1081	AATAGCCTGCGAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTGACTCCCAAGTTAAG	1140
QY	1184	AGTGAGTTTCAACAGCCTGTCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT	1243
Db	1141	AGTGAGTTTCAACAGCCTGTCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT	1200
QY	1244	CTTTCTGAAAGTTCTCTCTCCAAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC	1303
Db	1201	CTTTCTGAAAGTTCTCTCTCCAAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC	1260
QY	1304	TGCGTGGAAATGTCAGAAAGCAGTCTATCCAAATGGAGCGTCTCTTGGCCAACCCAGCAGGTG	1363
Db	1261	TGCGTGGAAATGTCAGAAAGCAGTCTATCCAAATGGAGCGTCTCTTGGCCAACCCAGCAGGTG	1320
QY	1364	TTTCACATCAGCTGCTTCCGTTGCTCTATTTGCAACAAACACTCAGTCTAGGAACATAT	1423
Db	1321	TTTCACATCAGCTGCTTCCGTTGCTCTATTTGCAACAAACACTCAGTCTAGGAACATAT	1380
QY	1424	GCATCTTTACATCGAAGAACTATTGTAAGCCTCATTCAATCAACTCTTTAAATCTAAG	1483
Db	1381	GCATCTTTACATCGAAGAACTATTGTAAGCCTCATTCAATCAACTCTTTAAATCTAAG	1440
QY	1484	GGCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAAAT	1543
Db	1441	GGCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAAAT	1500
QY	1544	GAAAAAGAAAGATTTTGGAGAGCAGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCAC	1603
Db	1501	GAAAAAGAAAGATTTTGGAGAGCAGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCAC	1560
QY	1604	AGCCAGGGGTAGAAAGATGCCCTATTGCTAAGTGGGTGTCCTGGCTCAAGTATGGAA	1663
Db	1561	AGCCAGGGGTAGAAAGATGCCCTATTGCTAAGTGGGTGTCCTGGCTCAAGTATGGAA	1620
QY	1664	GCCAAAGCCCTCTCTCAGCAGGAGGAGGAGACACAGCAGCTGAAACCAAGAGCTGAGG	1723
Db	1621	GCCAAAGCCCTCTCTCAGCAGGAGGAGGAGACACAGCAGCTGAAACCAAGAGCTGAGG	1680
QY	1724	ATCGCTGGCCACCCCTCACTGAACTTGGAAAGTTTCAAGGAGTGCCTTGGAGAAAGGATC	1783
Db	1681	ATCGCTGGCCACCCCTCACTGAACTTGGAAAGTTTCAAGGAGTGCCTTGGAGAAAGGATC	1740
QY	1784	AAATGTCAAAAGCCCAAAATGGCTCTGAAAGACGAAATCAGCAAGCCCGAAAGTTCTTGAG	1843
Db	1741	AAATGTCAAAAGCCCAAAATGGCTCTGAAAGACGAAATCAGCAAGCCCGAAAGTTCTTGAG	1800
QY	1844	GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGAAAGAGCCGC	1903
Db	1801	GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGAAAGAGCCGC	1860
QY	1904	CCATTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGCAAGAGCCCAAAACTCTGTCC	1963
Db	1861	CCATTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGCAAGAGCCCAAAACTCTGTCC	1920
QY	1964	CACTTATCAGGAAAGGCTTGAAGCATGTGAGCAGAGTGAAGAGTCTGTGGTGGAGAGA	2023
Db	1921	CACTTATCAGGAAAGGCTTGAAGCATGTGAGCAGAGTGAAGAGTCTGTGGTGGAGAGA	1980
QY	2024	GTTCAGAAAGGAAACAAAGTGGAAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGA	2083

Db 1981 ||||| GTTGCAAGGAAACAAAGTGGAAATGCCAAGCTTCTAAGAAGAAATGGGAATGTGGGA 2040
Qy 2084 ABAACACCTGGCAACAAAGAAATCTAAGAGAGACAGAGGGAAGAGTAAGTAAGGT 2143
Db 2041 ABAACAACTGGCAACAAAGAAATCTAAGAGAGACAGAGGGAAGAGTAAGTAAGGT 2100
Qy 2144 CATAGTTTGGAGATGGAGATGAGAAATCTGTAGAAAATGGTGCACACTCCGATGAAGAT 2203
Db 2101 CATAGTTTGGAGATGGAGATGAGAAATCTGTAGAAAATGGTGCACACTCCGATGAAGAT 2160
Qy 2204 GATAACAGCTTCTCAAAACAAATCTCCACAAAGAACCCCAAGTCTCTGAATTTGGTCCAGT 2263
Db 2161 GATAACAGCTTCTCAAAACAAATCTCCACAAAGAACCCCAAGTCTCTGAATTTGGTCCAGT 2220
Qy 2264 TTTGTAGACAAACCTTTTGTGTAAGAAATCTCACTACTACAGAAATCCCAAGATGTG 2323
Db 2221 TTTGTAGACAAACCTTTTGTGTAAGAAATCTCACTACTACAGAAATCCCAAGATGTG 2280
Qy 2324 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAAGAAACAGATAAAGAGAAAT 2383
Db 2281 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAAGAAACAGATAAAGAGAAAT 2340
Qy 2384 CGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATGCAATGATGCTGGCCCTTAAA 2443
Db 2341 CGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATGCAATGATGCTGGCCCTTAAA 2400
Qy 2444 TTCAATGTTAGTGTAGCGACCACTGCCCTTTGTCAAAATGTGATCCACATTAAGCAGGTA 2503
Db 2401 TTCAATGTTAGTGTAGCGACCACTGCCCTTTGTCAAAATGTGATCCACATTAAGCAGGTA 2460
Qy 2504 TCCAGCATGAATGTAAATTTACTTGGAGTAACCTTTGAAAAGAAATTCCTTCTTAAAT 2563
Db 2461 TCCAGCATGAATGTAAATTTACTTGGAGTAACCTTTGAAAAGAAATTCCTTCTTAAAT 2520
Qy 2564 CAAAAACAAACAAAAACAAAAACAAATCTTAATCTAGAGATAAATTTACTTTA 2623
Db 2521 CAAAAACAAACAAAAACAAAAACAAATCTTAATCTAGAGATAAATTTACTTTA 2580
Qy 2624 AATTCCTCAATTTAGCAGTGATGATATGATAGTCTGTAAAGCTTTGAACTGGGGAAA 2683
Db 2581 AATTCCTCAATTTAGCAGTGATGATATGATAGTCTGTAAAGCTTTGAACTGGGGAAA 2640
Qy 2684 TATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGTAGA 2743
Db 2641 TATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGTAGA 2700
Qy 2744 TAGATGATTAGTAGTATTTGTACACACTATTTTGGAAATTAAGAGACATACAGAGGAA 2803
Db 2701 TAGATGATTAGTAGTATTTGTACACACTATTTTGGAAATTAAGAGACATACAGAGGAA 2760
Qy 2804 TTTAGGGCTTAAACATTACGACTGAATGCATTTAGTATAAAGGCGACAGTTTGTATAT 2863
Db 2761 TTTAGGGCTTAAACATTACGACTGAATGCATTTAGTATAAAGGCGACAGTTTGTATAT 2820
Qy 2864 TTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGTAGCTT 2923
Db 2821 TTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGTAGCTT 2880
Qy 2924 TAAATTTTTAGTTAATTTTCTGCTGTATATATAGAGAAATTTACTATTTATGTC 2983
Db 2881 TAAATTTTTAGTTAATTTTCTGCTGTATATATAGAGAAATTTACTATTTATGTC 2940
Qy 2984 CTGCTCTCTAAACTACATCCTGAACCTCGACGCTCTGAGGTATAATACACAGAGACATTT 3043
Db 2941 CTGCTCTCTAAACTACATCCTGAACCTCGACGCTCTGAGGTATAATACACAGAGACATTT 3000
Qy 3044 TTGAGCAATTTGAAAACCAACCTACCTCTTTCGGTGTCTAGAGAGATCTGCTGCTCCC 3103
Db 3001 TTGAGCAATTTGAAAACCAACCTACCTCTTTCGGTGTCTAGAGAGATCTGCTGCTCCC 3060
Qy 3104 AATAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTCT 3163

3061 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTTCT 3120
Qy 3164 GGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3223
Db 3121 GGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3180
Qy 3224 ATCACTGTCCCCCTCTTCCGTTGTAGAGCAAGTGAAGAGCTTTAAAGAGGAGAGAA 3283
Db 3181 ATCACTGTCCCCCTCTTCCGTTGTAGAGCAAGTGAAGAGCTTTAAAGAGGAGAGAA 3240
Qy 3284 AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGC 3343
Db 3241 AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGC 3300
Qy 3344 CCTTTTGTAGACACTAATTTTAAATCTTACTAGCTCTGAAATATATTTGATTTTATC 3403
Db 3301 CCTTTTGTAGACACTAATTTTAAATCTTACTAGCTCTGAAATATATTTGATTTTATC 3360
Qy 3404 ACAGTATTTCTCAGGTTGAAATTAACCAACTATAGGCCCTTTTCTTGGGATGATTTCTA 3463
Db 3361 ACAGTATTTCTCAGGTTGAAATTAACCAACTATAGGCCCTTTTCTTGGGATGATTTCTA 3420
Qy 3464 GTCTTAAGGTTTGGGACATTTATAAAGCTTTAGTACATTTTGTGTACACAGTTGATTTCC 3523
Db 3421 GTCTTAAGGTTTGGGACATTTATAAAGCTTTAGTACATTTTGTGTACACAGTTGATTTCC 3480
Qy 3524 AAATTTGATGATGGAGGGAGAGGTGTCTTAAAGCTGTAGGCTTTTCTTGTACTGCAAT 3583
Db 3481 AAATTTGATGATGGAGGGAGAGGTGTCTTAAAGCTGTAGGCTTTTCTTGTACTGCAAT 3540
Qy 3584 TATAGAGATTTAGCTTTAATATTTTTTAGAGATGATAAACAATCTGCTTTCTTGTAGCTTA 3643
Db 3541 TATAGAGATTTAGCTTTAATATTTTTTAGAGATGATAAACAATCTGCTTTCTTGTAGCTTA 3600
Qy 3644 CCTAGCTGAAACATTTTATTTCAATTAAGATTTTAAATTAATAAATTTGAAAAA 3703
Db 3601 CCTAGCTGAAACATTTTATTTCAATTAAGATTTTAAATTAATAAATTTGAAAAA 3660
Qy 3704 AA 3705
Db 3661 AA 3662

RESULT 2
HSM800164
LOCUS HSM800164 3664 bp mRNA linear PRI 03-AUG-2001
DEFINITION Homo sapiens mRNA; cDNA DKFZp586i1918 (from clone DKFZp586i1918);
complete cds.
ACCESSION AL136911
VERSION AL136911.1 GI:12053316
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3664)
AUTHORS Wamburt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2001) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOZA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp586i1918) is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clonesrzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/
Location/Qualifiers
source
1. 3664
/organism="Homo sapiens"
/mol_type="mRNA"

```
/db_xref="taxon:9606"
/map="12"
/clone="DKF2p586i1918"
/tissue_type="uterus"
/clone_lib="586 (synonym: hutei). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1. 3664
/gene="DKF2p586i1918"
CDS
94..2373
/gene="DKF2p586i1918"
/note="similarity to transcription factor SF3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB56845.1"
/db_xref="GI:1203317"
/db_xref="SPTREMBL:Q5UHB6"
/translation="MESPFNRQWTSLSRVITAKLSLVNKNKSSAIVEIFSKYOKA
AEETNMBKRGNTENLQGLHFKGDTLVKKWENPGLGASHDLSLNSSTETRHAD
HPAEVTSHAASGAKAQDEOIHPRSLRSPPEALVOGRYPHIKDGEDLDKHTEKK
MENCLGSRHEVKESEISENTDASGKIEKYNPLNRLKMFKEGETQTKILRAQSR
ASGKRKSENSYSLDLIGPQLSSSTFDSEKNESRNLLELPLRSETISKRMAYQA
AVSKQSSNTYNLSKASGGEIKHKMEQKENVPPPEVCI THQGEKIKIADNENSLAV
RSTPAEDSDRSQVSKVQHPKPLSPDRASSLSSESPKMKKFOAPARETCVE
COKTVYMERLLAQOVPHISFCRSCYNKLSLGTYSLSLHRIYCKPHFNQLPKSG
NYDEGFGHRPHKDLWASKNEEILERPAQLANARETPHSPGVEDAPIAKVGLAASM
EAKASSQEREDKPAETPKRIAPPPTELGSSALUEGIKMSKPKWPPDEISKE
VPEDVDLKLRLSSLSKERSRPTVAASFQSTSVKSPKTVSPPIRKWMSGEQEE
SVGRVAERKOVENAKSKNGNVGKTTWQNKESGETKRSKESHSLENNENLVEN
GADSEDDNDFLKQSQPOEKSLNWSFVNDTFAEETTTQNKSQDVLEWEGEVVEL
SVEEQIKRNYDEDEEE"
3637, 3642
polyA_signal
/gene="DKF2p586i1918"
3647
polyA_site
/gene="DKF2p586i1918"
1210 a 734 c 809 g 911 t
BASE COUNT
ORIGIN
Query Match 98.8%; Score 3660.4; DB 9; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 44 GCGCTAGGTAGCGCGCGGACCTGTGACAGGGGTGTGACAGCGAGAGAAAGCGGC 103
Db 1 GCGCTAGGTAGCGCGCGGACCTGTGACAGGGGTGTGACAGCGAGAGAAAGCGGC 60
QY 104 TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTATATACGG 163
Db 61 TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTATATACGG 120
QY 164 CAATGGACCTCACTATCATTTAGGGGTACAGCCAAAGAACTTTCTCTGTGCAACAAGAAC 223
Db 121 CAATGGACCTCACTATCATTTAGGGGTAAACAGCCAAAGAACTTTCTCTGTGCAACAAGAAC 180
QY 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAC 283
Db 181 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAC 240
QY 284 ATGGAGAAGAAGAGAGTAACAACCGAAAATCTCTCCAGACACTTTAGAAAGGGGCCCTG 343
Db 241 ATGGAGAAGAAGAGAGTAACAACCGAAAATCTCTCCAGACACTTTAGAAAGGGGCCCTG 300
QY 344 ACTGTGTTAAAGAAAGTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCT 403
Db 301 ACTGTGTTAAAGAAAGTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCT 360
QY 404 CTACGGACAGCAGCAGTGAATAGGCACAGAGCAGACATCCTCTCTGCTGAAAGTGACA 463
Db 361 CTACGGACAGCAGCAGTGAATAGGCACAGAGCAGACATCCTCTCTGCTGAAAGTGACA 420
QY 464 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGACAAATCCACCCACAGACTCTAGA 523
Db 421 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGACAAATCCACCCACAGACTCTAGA 480
```

```
QY 524 CTCAGGTCACTCTCTGAAAGCCCTGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAG 583
Db 481 CTCAGGTCACTCTCTGAAAGCCCTGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAG 540
QY 584 GATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATTTGCTAGGAGAAATCCAGG 643
Db 541 GATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATTTGCTAGGAGAAATCCAGG 600
QY 644 CATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAA 703
Db 601 CATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAA 660
QY 704 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT 763
Db 661 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT 720
QY 764 AAGATTCTCGGGCCCCAAGCGCAAGTGAAGTGAAGGAAGATCTCTGAAACAGCTAT 823
Db 721 AAGATTCTCGGGCCCCAAGCGCAAGTGAAGTGAAGGAAGATCTCTGAAACAGCTAT 780
QY 824 TCTCTAGATGACTGTGAATAGGCCAGGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 883
Db 781 TCTCTAGATGACTGTGAATAGGCCAGGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 840
QY 884 AAAAAATGAGAGTAGACGAAATCTGGAATCTCCACGCTCTCAGAAAACCTCTATAAAGGAT 943
Db 841 AAAAAATGAGAGTAGACGAAATCTGGAATCTCCACGCTCTCAGAAAACCTCTATAAAGGAT 900
QY 944 CGAATGGCCCAAGTACAGGCGAGCTGTGTCAAAACAAAAGCAGCTCAACCAACTATACAAAT 1003
Db 901 CGAATGGCCCAAGTACAGGCGAGCTGTGTCAAAACAAAAGCAGCTCAACCAACTATACAAAT 960
QY 1004 GAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTG 1063
Db 961 GAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTG 1020
QY 1064 CCCCAGGTCCTGAGTCTGCAATCAACCATCAGGAAGGGGAAAGATTTCTGCAATAGAG 1123
Db 1021 CCCCAGGTCCTGAGTCTGCAATCAACCATCAGGAAGGGGAAAGATTTCTGCAATAGAG 1080
QY 1124 AATAGCCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCCGAGGTTAAG 1183
Db 1081 AATAGCCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCCGAGGTTAAG 1140
QY 1184 AGTAGAGTTCAACAGCTGTCCATCCCAAGCCATTAAGTCCAGATTCAGAGCCTCCAGT 1243
Db 1141 AGTAGAGTTCAACAGCTGTCCATCCCAAGCCATTAAGTCCAGATTCAGAGCCTCCAGT 1200
QY 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAGACC 1303
Db 1201 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAGACC 1260
QY 1304 TCGGTGGAATGTGAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCCAACCCAGCAGTG 1363
Db 1261 TCGGTGGAATGTGAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCCAACCCAGCAGTG 1320
QY 1364 TTTTCACTAGCTGCTCCGTTGCTCTCTATTTGCAACCAAACTCAGTCTAGGAAACATAT 1423
Db 1321 TTTTCACTAGCTGCTCCGTTGCTCTCTATTTGCAACCAAACTCAGTCTAGGAAACATAT 1380
QY 1424 GCATCTTTAGATGAAGAAATCTATTGAAGCCTCACTTCAATCAACTCTTTTAAATCTTAAG 1483
Db 1381 GCATCTTTAGATGAAGAAATCTATTGAAGCCTCACTTCAATCAACTCTTTTAAATCTTAAG 1440
QY 1484 GGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAAT 1543
Db 1441 GGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAAT 1500
QY 1544 GAAAAACGAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAAATGCAAGGGAGAGACCCCTCAC 1603
Db 1501 GAAAAACGAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAAATGCAAGGGAGAGACCCCTCAC 1560
```

QY	1604	AGCCAGGGGTAGAAAGATGCCCTATTGCTAAGTGGGTCTCTGCTGCAAGTATGGAA	1663
Db	1561	AGCCAGGGGTAGAAAGATGCCCTATTGCTAAGTGGGTCTCTGCTGCAAGTATGGAA	1620
QY	1664	GCCAAAGCCTCTCTCAGCAGGAGAGAGACAAGCCAGCTGAAACCAAGAAAGCTGAGG	1723
Db	1621	GCCAAAGCCTCTCTCTCAGCAGGAGAGAGACAAGCCAGCTGAAACCAAGAAAGCTGAGG	1680
QY	1724	ATCGCTGGCCACCCCCCACTGAACTTTGGAAGTTTCAAGAGTGCCTTTGAGGAAGGGATC	1783
Db	1681	ATCGCTGGCCACCCCCCACTGAACTTTGGAAGTTTCAAGAGTGCCTTTGAGGAAGGGATC	1740
QY	1784	AAATGTCAAAGCCCAATAGCCCTCTGTAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAG	1843
Db	1741	AAATGTCAAAGCCCAATAGCCCTCTGTAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAG	1800
QY	1844	GATGTGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGCCGC	1903
Db	1801	GATGTGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGCCGC	1860
QY	1904	CCATTCACTGTACAGCTTCATTTCAAAGCACCTCTGTCAAAGGCCCAAAAACCTGTGTC	1963
Db	1861	CCATTCACTGTACAGCTTCATTTCAAAGCACCTCTGTCAAAGGCCCAAAAACCTGTGTC	1920
QY	1964	CCACCTATCAGGAAGGCTGGAGCATGTTCAGACAGAGTGAAGAGTCTGTGGTGGAGA	2023
Db	1921	CCACCTATCAGGAAGGCTGGAGCATGTTCAGACAGAGTGAAGAGTCTGTGGTGGAGA	1980
QY	2024	GTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTTAAGAAATGGGAATGTGGGA	2083
Db	1981	GTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTTAAGAAATGGGAATGTGGGA	2040
QY	2084	AAAAACAACCTGGCAAAACAAGAAATCTAAGAGAGACAGACAGGAAGAGAAAGT	2143
Db	2041	AAAAACAACCTGGCAAAACAAGAAATCTAAGAGAGACAGACAGGAAGAGAAAGT	2100
QY	2144	CATAGTTTGGAGATGAGAAATGAGAAATCTTTAGAAAATGGTGACAGCTCCGATCAAGAT	2203
Db	2101	CATAGTTTGGAGATGAGAAATGAGAAATCTTTAGAAAATGGTGACAGCTCCGATCAAGAT	2160
QY	2204	GATAACAGCTTCTCAAACAACAATCTCCAAGAAACCCAACTCTCTGAATTTGGTCGAGT	2263
Db	2161	GATAACAGCTTCTCAAACAACAATCTCCAAGAAACCCAACTCTCTGAATTTGGTCGAGT	2220
QY	2264	TTTGTAGACAAACACCTTTGCTGAAGAAATTCACACTCAGAAATCCAGAAATCCAGGATGTG	2323
Db	2221	TTTGTAGACAAACACCTTTGCTGAAGAAATTCACACTCAGAAATCCAGAAATCCAGGATGTG	2280
QY	2324	GAACTCTGGAGGGAGAGTGTGTCAAAGAGCTCTCTGTGGAAGACAGATTAAGAGAAAT	2383
Db	2281	GAACTCTGGAGGGAGAGTGTGTCAAAGAGCTCTCTGTGGAAGACAGATTAAGAGAAAT	2340
QY	2384	CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTAAA	2443
Db	2341	CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTAAA	2400
QY	2444	TTTCATGTTAGTGTAGCAGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTA	2503
Db	2401	TTTCATGTTAGTGTAGCAGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTA	2460
QY	2504	TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAAAGAAATTCCTTTTAAAT	2563
Db	2461	TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAAAGAAATTCCTTTTAAAT	2520
QY	2564	CAAAAAACAAAAACAAAAACAAAAACAACTTCTAAATACTAGAGATACCTTTACTTTA	2623
Db	2521	CAAAAAACAAAAACAAAAACAAAAACAACTTCTAAATACTAGAGATACCTTTACTTTA	2580
QY	2624	AATTTCTTCAATTTAGCAGTGATGATGCAATAGTGTGTAAGGCTTGTAACTGGGAAA	2683
Db	2581	AATTTCTTCAATTTAGCAGTGATGATGCAATAGTGTGTAAGGCTTGTAACTGGGAAA	2640
QY	2684	TATCCACCTGATAATAGCCAGATTTCTACTGTATTTCCAAAAAGGCAATATTAAGGTAGA	2743

RESULT 3

Db	2641	TATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCAAAAAGCAATATTAAGGTAGA	2700
QY	2744	TAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAGAA	2803
Db	2701	TAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAGAA	2760
QY	2804	TTTAGGGCTTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATAT	2863
Db	2761	TTTAGGGCTTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATAT	2820
QY	2864	TTTTAAATGAATACCAATTTAAATTTTTTAGTATTTACCTGTTAAAGAGATTTATTTAGTCTT	2923
Db	2821	TTTTAAATGAATACCAATTTAAATTTTTTAGTATTTACCTGTTAAAGAGATTTATTTAGTCTT	2880
QY	2924	TAAATTTTTTAGGTTAAATTTCTTTGCTGTGATATATATGAGGAATTTACTACTTTATGTC	2983
Db	2881	TAAATTTTTTAGGTTAAATTTCTTTGCTGTGATATATATGAGGAATTTACTACTTTATGTC	2940
QY	2984	CTGCTCTCTAAACTACATCTCGAACTCGACCTCTCGAGGTATAATACAAAGAGACACTTT	3043
Db	2941	CTGCTCTCTAAACTACATCTCGAACTCGACCTCTCGAGGTATAATACAAAGAGACACTTT	3000
QY	3044	TTGAGGCAATTTGAAAAACCAACCTACACTCTTCGGTGTCTAGAGAGATCTGCTCTCC	3103
Db	3001	TTGAGGCAATTTGAAAAACCAACCTACACTCTTCGGTGTCTAGAGAGATCTGCTCTCC	3060
QY	3104	AAATAAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTCT	3163
Db	3061	AAATAAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTCT	3120
QY	3164	GGTGATATCTGTGCTTCTCATTAATTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG	3223
Db	3121	GGTGATATCTGTGCTTCTCATTAATTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG	3180
QY	3224	ATCACTGTCCCCCATCTTCGGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAAAGAA	3283
Db	3181	ATCACTGTCCCCCATCTTCGGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAAAGAA	3240
QY	3284	AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCTGTATTTCCCTTATGATGTC	3343
Db	3241	AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCTGTATTTCCCTTATGATGTC	3300
QY	3344	CCCTTTTGTAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATC	3403
Db	3301	CCCTTTTGTAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATC	3360
QY	3404	ACAGTATTCTCAGGGTGAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTTCTA	3463
Db	3361	ACAGTATTCTCAGGGTGAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTTCTA	3420
QY	3464	GTCTTAAGTTTGGGGACATTTAAACTTGAGTACATTTGTTGTACACAGTTGATATTC	3523
Db	3421	GTCTTAAGTTTGGGGACATTTAAACTTGAGTACATTTGTTGTACACAGTTGATATTC	3480
QY	3524	AAATTGTATGAGTGGGAGGAGAGGTGCTTAAGCTGTAGGCTTTCTTTGTACTGCAAT	3583
Db	3481	AAATTGTATGAGTGGGAGGAGAGGTGCTTAAGCTGTAGGCTTTCTTTGTACTGCAAT	3540
QY	3584	TATAGAGATTTAGCTTTAAATTTTTTAGAGATGTAAGACATTTCTGCTTTCTTAGTCTTA	3643
Db	3541	TATAGAGATTTAGCTTTAAATTTTTTAGAGATGTAAGACATTTCTGCTTTCTTAGTCTTA	3600
QY	3644	CCTAGTCTGAACATTTTATTCATTAAGATTTTAAATTTAAATTTGAAAAACAAAAA	3703
Db	3601	CCTAGTCTGAACATTTTATTCATTAAGATTTTAAATTTAAATTTGAAAAACAAAAA	3660
QY	3704	AA 3705	
Db	3661	AA 3662	

AF198454
LOCUS AF198454 3655 bp mRNA linear PRI 10-JAN-2000
DEFINITION Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA, complete cds.
ACCESSION AF198454
VERSION AF198454.1 GI:6685006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
JOURNAL EPLIN, epithelial protein lost in neoplasm
MEDLINE Oncogene 18 (54), 7838-7841 (1999)
PUBMED 20087188
REFERENCE 2 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
JOURNAL Direct Submission
TITLE Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
FEATURES
source
1...3655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
1...3655
/gene="EPLIN"
102...2381
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAF23755.1"
/db_xref="GI:6685007"
translation="MESSPFNRRQWTSLSRVTAKELSLVNKNKSSAIVEIFSQYKA
AEETNMEKRSNTENLSOHRPKGLTLVKKKWPGLGASHDLSLRNSSTETRRHAD
HPPEAVTSHAASGAKADQEQIHPRSPLRPPALVQGRYPHIKDGEDLKDHSTESKK
MENCLOGSRHEVESSEISENTDASGTEKYNVPLNLKMFKEGEBTQIKLRAQYS
ASGRKISNYSUDDLGFGQJSSFTDSERKESRNLDELPLSLSETSIKDRNAKYQA
AVSKQSSNTYLLKASGEIKIHKMEKENVPQGEVCIHQEKEKISANENSLAV
RSTPAEDSDSDSQVQVHPKPLSPDSRASLSSESPPKAMKFOAPARETCVE
CKTVYPMERLLANQVQVHISCFRCYSNNKLSLGTYSALHGIYCKPHNOLFQSKG
NYDEGFCGRPHKDLWASKNNEBILERPQOLANARETPSPGVEDAPIKAVGLASM
EKAASSQEREDKPAETKRLRIAPPPTELGSSGSALEBGTIKMSKPWPPEDEISKE
VPEDVDLKLKSSLSKERSRPFTVAASFQTSVKSPKTVSPPIRKGMSSEQBE
SVGRVAERQVENAKSKNGVKTWNQKESKGETKRSKEGHSLEWENENLVEN
GASDEDDNSFLKQSQPOEPKSLNWSFVNDTFAEBFTTQNKQSQDVLEWEGEVWKE
SVREOIKNRYVDEDEDEE"
polyA_signal 3632..3637
/gene="EPLIN"
BASE COUNT 1197 a 737 c 810 g 911 t
ORIGIN
Query Match 98.3%; Score 3643.8; DB 9; Length 3655;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3645; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 44 GCGTAGGTAGAGCGCGGACCTGTGACAGGGGTGTGACAGCGGACAGAGAAAGCGGC 103
Db 9 GCGTAGGTAGAGCGCGGACCTGTGACAGGGGTGTGACAGCGGACAGAGAAAGCGGC 68
Qy 104 TTTTAGCAGGTATTTTCAAGTGTCTGTAGAGAAAGATGATCTCCATTAATAGACGG 163
Db 69 TTTTAGCAGGTATTTTCAAGTGTCTGTAGAGAAAGATGATCTCCATTAATAGACGG 128
Qy 164 CAATGGACCTCATTATTCAGGGTACAGCAAGAACTTTCTCTGTCAACAGAAC 223
Db 129 CAATGGACCTCATTATTCAGGGTACAGCAAGAACTTTCTCTGTCAACAGAAC 188
Qy 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAAGTACCAGAAAGCAGCTGAAGAAACAAAC 293

189 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 248
Qy 284 ATGGAGAGAAAGAGAGTAACACCCAAAATCTCTCCAGCAGCTTTAGAAAGGGGACCCCTG 343
Db 249 ATGGAGAGAAAGAGAGTAACACCCAAAATCTCTCCAGCAGCTTTAGAAAGGGGACCCCTG 308
Qy 344 ACTGTGTTAAAGAGAAAGTGGAGAAACCCAGGGCTGGGAGCAGAGTCTCTACACAGACTCT 403
Db 309 ACTGTGTTAAAGAGAAAGTGGAGAAACCCAGGGCTGGGAGCAGAGTCTCTACACAGACTCT 368
Qy 404 CTACGGAACAGCAGCACTGAGATTAGGCACAGAGCAGACCATCTCTCTGCTGAAGTGACA 463
Db 369 CTACGGAACAGCAGCACTGAGATTAGGCACAGAGCAGACCATCTCTCTGCTGAAGTGACA 428
Qy 464 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGACTCTAGA 523
Db 429 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGACTCTAGA 488
Qy 524 CTCAGGTCACCTCCTGAAGCCCTCGTTTCAGGGTCGATATCCCCACATCAAGGACCGTGAG 583
Db 489 CTCAGGTCACCTCCTGAAGCCCTCGTTTCAGGGTCGATATCCCCACATCAAGGACCGTGAG 548
Qy 584 GATCTTAAAGACCACTCAACAGAAAAGTAAAAAATGGAAAATTTGTCTAGGAAATCCAGG 643
Db 549 GATCTTAAAGACCACTCAACAGAAAAGTAAAAAATGGAAAATTTGTCTAGGAAATCCAGG 608
Qy 644 CATGAAGTAGAAAAATCAGAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAA 703
Db 609 CATGAAGTAGAAAAATCAGAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAA 668
Qy 704 TATAATGTTCCGCTCAACAGAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACT 763
Db 669 TATAATGTTCCGCTCAACAGAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACT 728
Qy 764 AAGATTCTCCGGGCCAAAGCCGAAAGTCAAGTGGAGGAAGAGATCTCTGAAAACAGCTAT 823
Db 729 AAGATTCTCCGGGCCAAAGCCGAAAGTCAAGTGGAGGAAGAGATCTCTGAAAACAGCTAT 788
Qy 824 TCTCTAGATGACTGGAAATAGGCCAGGTGAGTGTGATCTTCTACATTTGACTTCGGAG 883
Db 789 TCTCTAGATGACTGGAAATAGGCCAGGTGAGTGTGATCTTCTACATTTGACTTCGGAG 848
Qy 884 AAAAAATGAGTAGACGAGAAATCTGGAATCTCCAGCTTCACGCCCTCTCAGAAACCTCTATNAAGAT 943
Db 849 AAAAAATGAGTAGACGAGAAATCTGGAATCTCCAGCTTCACGCCCTCTCAGAAACCTCTATNAAGAT 908
Qy 944 CGAAATGGCCAAAGTACAGGCGAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAAT 1003
Db 909 CGAAATGGCCAAAGTACAGGCGAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAAT 968
Qy 1004 GAGCTGAAAGCCAGTGGTGGGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTG 1063
Db 969 GAGCTGAAAGCCAGTGGTGGGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTG 1028
Qy 1064 CCCCAGGTCCTGAGGTCTGATCAACCATCAGGAGGGGAAAGATTTCTGCAAAATAG 1123
Db 1029 CCCCAGGTCCTGAGGTCTGATCAACCATCAGGAGGGGAAAGATTTCTGCAAAATAG 1088
Qy 1124 AATAGCCTGGAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTACTCCAGGTTAAG 1183
Db 1089 AATAGCCTGGAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTACTCCAGGTTAAG 1148
Qy 1184 AGTAGAGTTCAACAGCCTGTCCATCCCAAGCCACTTAAGTCCAGATTCCAGAGCCTCCAGT 1243
Db 1149 AGTAGAGTTCAACAGCCTGTCCATCCCAAGCCACTTAAGTCCAGATTCCAGAGCCTCCAGT 1208
Qy 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1303
Db 1209 CTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1268
Qy 1304 TSCGTGGAATGTGAGAGACAGTCTATCCCAATGGAGCGTCTCTTGGCCAAACCCAGCGTG 1363

D	b	1269	TGCGTGGAAATGTCAAGAGACAGTCTTATCCAATTGGAGCGTCTCTTTGGCCAAACCAGCAGGTG	1320
Q	y	1364	TTTTACATCAGCTGCTTCGGTTGCTCCTATTGCACACAACAACCTCAGTCTTAGGAACTAT	1423
D	b	1329	TTTTACATCAGCTGCTTCGGTTGCTCCTATTGCACACAACAACCTCAGTCTTAGGAACATAT	1388
Q	y	1424	GCATCTTTACATGGAGAAGATCTATTGTAAGCCTCAC TTCATCAACTCTTTTAAATCTFAAG	1483
D	b	1389	GCATCTTTACATGGAGAAGATCTATTGTAAGCCTCAC TTCATCAACTCTTTTAAATCTFAAG	1448
Q	y	1484	GCCAACTATGATGAAGGCTTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAAAAT	1543
D	b	1449	GGCAACTATGATGAAGGCTTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAAAAT	1508
Q	y	1544	GAAAAACGAAGAGATTTTGGAGAGACAGCCCAGCTTGCAATGCAAGGAGACCCCTCAC	1603
D	b	1509	GAAAAACGAAGAGATTTTGGAGAGACAGCCCAGCTTGCAATGCAAGGAGACCCCTCAC	1568
Q	y	1604	AGCCCCAGGGGTAGAAGATGCCCTTATGCTAAGCTGGGTCTCTGGCTGCAAGTATGGAA	1663
D	b	1569	AGCCCCAGGGGTAGAAGATGCCCTTATGCTAAGCTGGGTCTCTGGCTGCAAGTATGGAA	1628
Q	y	1664	GCCAAAGGCTCTCTTCAGCAGGAGAAAGAACAGCAGCTGAAACCAAGAAGCTGAGG	1723
D	b	1629	GCCAAAGGCTCTCTTCAGCAGGAGAAAGAACAGCAGCTGAAACCAAGAAGCTGAGG	1688
Q	y	1724	ATCGCTGGCCACCOCCECACTGAAC TTTGGAAGTTTCAGAAAGTCCCTTGGAGGAAGGATC	1783
D	b	1689	ATCGCTGGCCACCOCCECACTGAAC TTTGGAAGTTTCAGAAAGTCCCTTGGAGGAAGGATC	1748
Q	y	1784	AAATGTCAAAGCCCAATAGCCTCTCTGAAGACGAAATCAGAACGCCAGGTTCCCTGAG	1843
D	b	1749	AAATGTCAAAGCCCAATAGCCTCTCTGAAGACGAAATCAGAACGCCAGGTTCCCTGAG	1808
Q	y	1844	GATGTGCAATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAAGACCCGC	1903
D	b	1809	GATGTGCAATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAAGACCCGC	1868
Q	y	1904	CCATTCACTGTAGCAGCTTCATTTC AAAGCACCTCTGTC AAGAGCCCAAAACTGTGTCC	1963
D	b	1869	CCATTCACTGTAGCAGCTTCATTTC AAAGCACCTCTGTC AAGAGCCCAAAACTGTGTCC	1928
Q	y	1964	CCACTATCAGGAAAGGCTGGAGCATGT CAGAGCAGAGTGAAGAGTCTGTGGTGGAAAG	2023
D	b	1929	CCACTATCAGGAAAGGCTGGAGCATGT CAGAGCAGAGTGAAGAGTCTGTGGTGGAAAG	1988
Q	y	2024	GTTGCGAAGGAAACAAAGTGGAAATTC CAAGGCTTCTAAGSAGAAATGGATGTGGG	2083
D	b	1989	GTTGCGAAGGAAACAAAGTGGAAATTC CAAGGCTTCTAAGSAGAAATGGATGTGGG	2048
Q	y	2084	AAPACACCTGGCAAAACAAAGATCTAA AGGAGAGACAGGGAAGAGAGTGAAGAGGT	2143
D	b	2049	AAAACACCTGGCAAAACAAAGATCTAA AGGAGAGACAGGGAAGAGAGTGAAGAGGT	2108
Q	y	2144	CATAGTTTGGAGATGGAGAATGAGAATCTTGTAGAAAATGGTGCAGACTCCGATGAAGAT	2203
D	b	2109	CATAGTTTGGAGATGGAGAATGAGAATCTTGTAGAAAATGGTGCAGACTCCGATGAAGAT	2168
Q	y	2204	GATAACAGCTTCCTCAAACCAACATCTCCA CAGAACCCTCAAGTCTCTGAAATTTGGTCAGT	2263
D	b	2169	GATAACAGCTTCCTCAAACCAACATCTCCA CAGAACCCTCAAGTCTCTGAAATTTGGTCAGT	2228
Q	y	2264	TTTGTAGAACACCTTTTGTCTGAAGAAATTC ACTACTCAGAATCAGAAATCCCAGGATGTG	2323
D	b	2229	TTTGTAGAACACCTTTTGTCTGAAGAAATTC ACTACTCAGAATCAGAAATCCCAGGATGTG	2288
Q	y	2324	GAACCTCGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAGAGACAGATAAAGAGAAAT	2383
D	b	2289	GAACCTCGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAGAGAGACAGATAAAGAGAAAT	2348
Q	y	2384	CGGTATTATCATCAGGATGAGGATGAAGAGTGACAAATTC GAATGCTGGGCCCTTAAA	2443
D	b	2349	CGGTATTATCATCAGGATGAGGATGAAGAGTGACAAATTC GAATGCTGGGCCCTTAAA	2408

QY	2444	TT	CATGTTTAGTGT	TAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAGGTA	2503
DB	2409	TT	CATGTTTAGTGT	TAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAGGTA	2468
QY	2504	TCC	CAGCATGAAATGT	TAATTTACTCTTGGAAAGTAACCTTTTGGAAAAAAATTCCTCTTTAAAAAT	2563
DB	2469	TCC	CAGCATGAAATGT	TAATTTACTCTTGGAAAGTAACCTTTTGGAAAAAAATTCCTCTTTAAAAAT	2528
QY	2564	CA	AAAAACAAAAACAAAAACAAAAACAAAAACACATCTTAAAT	TACTAGAGATAAATCTTTACTTTA	2623
DB	2529	CA	AAAAACAAAAACAAAAACAAAAACAAAAACACATCTTAAAT	TACTAGAGATAAATCTTTACTTTA	2588
QY	2624	AA	TCTTCATTTTAGCAGT	GATGATGCAATAAGTGTCTGAAGCCTTGTAACTGGGGAAA	2683
DB	2589	AA	TCTTCATTTTAGCAGT	GATGATGCAATAAGTGTCTGAAGCCTTGTAACTGGGGAAA	2648
QY	2684	TAT	TCCACCTGAT	TAATAGCCAGATTTCTACTGTATTTCCAAAAAGGCAATATTAAGGTAGA	2743
DB	2649	TAT	TCCACCTGAT	TAATAGCCAGATTTCTACTGTATTTCCAAAAAGGCAATATTAAGGTAGA	2708
QY	2744	TAG	ATGATTAGTAGT	ATATTTGTTTACACACTATTTTGGAAATTAGAGAACATACACAGAAGGAA	2803
DB	2709	TAG	ATGATTAGTAGT	ATATTTGTTTACACACTATTTTGGAAATTAGAGAACATACACAGAAGGAA	2768
QY	2804	TT	TAGGGGCTTAAACAT	TAGCACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATAT	2863
DB	2769	TT	TAGGGGCTTAAACAT	TAGCACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATAT	2828
QY	2864	TTTT	TAATAGAAATACCAATTTAA	TTTTTTTAGTATTTTACCTGTTTAAGAGATATTTTAGTCTTT	2923
DB	2829	TTTT	TAATAGAAATACCAATTTAA	TTTTTTTAGTATTTTACCTGTTTAAGAGATATTTTAGTCTTT	2888
QY	2924	TAA	ATTTTTTAGGTTAAATTTTCT	TGCTGTGATATATAGAGAAATTTACTCTTTATGTC	2983
DB	2889	TAA	ATTTTTTAGGTTAAATTTTCT	TGCTGTGATATATAGAGAAATTTACTCTTTATGTC	2948
QY	2984	CTG	CTCTCTAAA	CTACATCTGAACTCGAGCTCCTCAGGTATAATACACACAGACACTTTT	3043
DB	2949	CTG	CTCTCTAAA	CTACATCTGAACTCGAGCTCCTCAGGTATAATACACACAGACACTTTT	3008
QY	3044	TT	GAGGCAATTGAAAAACCA	ACCTACACTCTTCGGTGTCTAGAGAGATCTGCTGTCTCCC	3103
DB	3009	TT	GAGGCAATTGAAAAACCA	ACCTACACTCTTCGGTGTCTAGAGAGATCTGCTGTCTCCC	3068
QY	3104	AA	ATAAGCTTTTGATCTGCC	AGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCT	3163
DB	3069	AA	ATAAGCTTTTGATCTGCC	AGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCT	3128
QY	3164	GGT	GATATCTGTCTCTCAT	TAATTAAGTAAAGCTGCAATATTTAGTAAATACCTTCGGG	3223
DB	3129	GGT	GATATCTGTCTCTCAT	TAATTAAGTAAAGCTGCAATATTTAGTAAATACCTTCGGG	3188
QY	3224	AT	CAGTGTCCCCCATCTTCC	GTGTAGAGCAAAGTGAAGAGTTTAAAGGAGGAAGAAGAA	3283
DB	3189	AT	CAGTGTCCCCCATCTTCC	GTGTAGAGCAAAGTGAAGAGTTTAAAGGAGGAAGAAGAA	3248
QY	3284	AGA	ACTGTCTTACACACACT	TGAGCTCAGACCTCTAAAACCTGTATTTCCCTTATGATGTC	3343
DB	3249	AGA	ACTGTCTTACACACACT	TGAGCTCAGACCTCTAAAACCTGTATTTCCCTTATGATGTC	3308
QY	3344	CC	CTTTTTCAGACACTAA	TTTTTAAATCTTACTAGCTCTGAAATATATTCGATTTTATC	3403
DB	3309	CC	CTTTTTCAGACACTAA	TTTTTAAATCTTACTAGCTCTGAAATATATTCGATTTTATC	3368
QY	3404	AC	AGTATTTCTCAGGGTGA	AAATTTAAACCAACTATAGGCCCTTTTCTCTGGGATGATTTTCTA	3463
DB	3369	AC	AGTATTTCTCAGGGTGA	AAATTTAAACCAACTATAGGCCCTTTTCTCTGGGATGATTTTCTA	3428
QY	3464	GT	CTTAAAGCTTTGGGACA	TATTAACCTTGAGTACATTTGTGTACACAGTTGATATTC	3523
DB	3429	GT	CTTAAAGCTTTGGGACA	TATTAACCTTGAGTACATTTGTGTACACAGTTGATATTC	3488

QY	1439	AGAATCTATTGTAGGCTCACTTCAATCAACTCTTTAAATCTTAAGGCGCAACTATGATGA	1498
Db	1299	AGAATCTATTGTAGGCTCACTTCAATCAACTCTTTAAATCTTAAGGCGCAACTATGATGA	1358
QY	1499	GGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAATGAAACCGAAGAGATT	1558
Db	1359	GGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAATGAAACCGAAGAGATT	1418
QY	1559	TTGGAGAGACCCAGCTTGCATATGCAAGGAGACCCCTCACAGCCAGGGGTAGAA	1618
Db	1419	TTGGAGAGACCCAGCTTGCATATGCAAGGAGACCCCTCACAGCCAGGGGTAGAA	1478
QY	1619	GATGCCCCCTATTCTAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCCCTCTCT	1678
Db	1479	GATGCCCCCTATTCTAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCCCTCTCT	1538
QY	1679	CAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCC	1738
Db	1539	CAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCC	1598
QY	1739	CCCACTGAACCTTGGAGTTTCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCC	1798
Db	1599	CCCACTGAACCTTGGAGTTTCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCC	1658
QY	1799	AAATGGCTCTTGAAGCAAAATCAGCAAGCCGCAAGTTCCCTGAGGATGTCATCTAGAT	1858
Db	1659	AAATGGCTCTTGAAGCAAAATCAGCAAGCCGCAAGTTCCCTGAGGATGTCATCTAGAT	1718
QY	1859	CTGAAGAAAGTAAAGACGATCTTCTTCACTGAAGGAAGAGCGCCCATCTCACTGTAGCA	1918
Db	1719	CTGAAGAAAGTAAAGACGATCTTCTTCACTGAAGGAAGAGCGCCCATCTCACTGTAGCA	1778
QY	1919	GCTTCAATTTCAAGCACTCTGTCAAGACCCCAAAACTGTGTCCCACTATCAGGAAA	1978
Db	1779	GCTTCAATTTCAAGCACTCTGTCAAGACCCCAAAACTGTGTCCCACTATCAGGAAA	1838
QY	1979	GGCTGAGATGTCAGACAGAGTGAAGAGTCTGTGGTGAAGAGTTGCAGAAAGGAAA	2038
Db	1839	GGCTGAGATGTCAGACAGAGTGAAGAGTCTGTGGTGAAGAGTTGCAGAAAGGAAA	1898
QY	2039	CAAGTGGAAAAATGCCAAGGCTTCTAAGAGAAATGGAAATGTGGAAAAACACCTGGCAA	2098
Db	1899	CAAGTGGAAAAATGCCAAGGCTTCTAAGAGAAATGGAAATGTGGAAAAACACCTGGCAA	1958
QY	2099	AACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAATGAAGGAAGTCAATAGTTTGAGATG	2158
Db	1959	AACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAATGAAGGAAGTCAATAGTTTGAGATG	2018
QY	2159	GAGAATGAGAATCTTGTAGAAAATGGTGCAACTCCGATGAAGATGATAACAGCTTCCTC	2218
Db	2019	GAGAATGAGAATCTTGTAGAAAATGGTGCAACTCCGATGAAGATGATAACAGCTTCCTC	2078
QY	2219	AAACAAATCTCCCAAGAACCCAGTCTCTGAATTTGTCAGTTTGTAGACAAACACC	2278
Db	2079	AAACAAATCTCCCAAGAACCCAGTCTCTGAATTTGTCAGTTTGTAGACAAACACC	2138
QY	2279	TTTGTCTGAAGAAATCTCACTACTCAGAAATCCAGGATGTGGAATCTCTGGAGGGA	2338
Db	2139	TTTGTCTGAAGAAATCTCACTACTCAGAAATCCAGGATGTGGAATCTCTGGAGGGA	2198
QY	2339	GAAGTGGTCAAGAGAGTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG	2398
Db	2199	GAAGTGGTCAAGAGAGTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG	2258
QY	2399	GATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTTA	2458
Db	2259	GATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTTA	2318
QY	2459	GGAGGCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTATCCAGCATGAATG	2518
Db	2319	GGAGGCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTATCCAGCATGAATG	2378

QY	2519	TAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTCTTAAATCAAAAAACAACAAA	2578
Db	2379	TAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTCTTAAATCAAAAAACAACAAA	2438
QY	2579	AAAAACAACAAAAACACACTTCTAAATACTAGAGATAAATTTACTTTACTTTCTTTAG	2638
Db	2439	AAAAACAACAAAAACACACTTCTAAATACTAGAGATAAATTTACTTTACTTTCTTTAG	2498
QY	2639	CAGTGATGATATGCAATATTTTGGAAATAGAGAACTACACAGAGGAAATTTAGGGCTTAAAC	2698
Db	2499	CAGTGATGATATGCAATATTTTGGAAATAGAGAACTACACAGAGGAAATTTAGGGCTTAAAC	2558
QY	2699	TAGCCCAATCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATAGTAGT	2758
Db	2559	TAGCCCAATCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATAGTAGT	2618
QY	2759	ATATTGTTTACACACTATTTTGGAAATAGAGAACTACACAGAGGAAATTTAGGGCTTAAAC	2818
Db	2619	ATATTGTTTACACACTATTTTGGAAATAGAGAACTACACAGAGGAAATTTAGGGCTTAAAC	2678
QY	2819	ATTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACC	2878
Db	2679	ATTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACC	2738
QY	2879	AATTTAATTTTTTGTATTTTACCTGTTAAGAGATATTTAGTCTTTTAAATTTTTTAGGTT	2938
Db	2739	AATTTAATTTTTTGTATTTTACCTGTTAAGAGATATTTAGTCTTTTAAATTTTTTAGGTT	2798
QY	2939	AATTTCTTGTGTGATATATAGAGAAATTTACTACTTTTATGTCCTCTCTAAACTA	2998
Db	2799	AATTTCTTGTGTGATATATAGAGAAATTTACTACTTTTATGTCCTCTCTAAACTA	2858
QY	2999	CATCTGAACCTCGACGCTCTCAGGTATAATAACAACAGAGCACTTTTGGAGCAATTTGAAA	3058
Db	2859	CATCTGAACCTCGACGCTCTCAGGTATAATAACAACAGAGCACTTTTGGAGCAATTTGAAA	2918
QY	3059	AACCAACCTACACTCTTTCGGTGCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGT	3118
Db	2919	AACCAACCTACACTCTTTCGGTGCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGT	2978
QY	3119	TCTGCCAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCTGGTGATATCTGTCT	3178
Db	2979	TCTGCCAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCTGGTGATATCTGTCT	3038
QY	3179	TCTCATATTTACTGAAAGCTGCATATTTTAGTAAATACCTTCGGGATCAGTGTCCCCAT	3238
Db	3039	TCTCATATTTACTGAAAGCTGCATATTTTAGTAAATACCTTCGGGATCAGTGTCCCCAT	3098
QY	3239	CTTCCGTGTTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTACAC	3298
Db	3099	CTTCCGTGTTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTACAC	3158
QY	3299	CACCTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGATGTCCCTTTTGGAGACAC	3358
Db	3159	CACCTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGATGTCCCTTTTGGAGACAC	3318
QY	3359	TAATTTTTAAATCTACTAGCTCTGAAATATATTTGATTTTTTATCAGATTTCTCAGGG	3418
Db	3219	TAATTTTTAAATCTACTAGCTCTGAAATATATTTGATTTTTTATCAGATTTCTCAGGG	3278
QY	3419	TGAATTTAAACCAACTATAGCCCTTTTCTTGGATGATTTTCTAGTCTTTAAGGTTTGGG	3478
Db	3279	TGAATTTAAACCAACTATAGCCCTTTTCTTGGATGATTTTCTAGTCTTTAAGGTTTGGG	3338
QY	3479	GACATTATAAAGTGTAGTACATTTTGTACAGATTTGATATTTCCAAATTTGATGGATGG	3538
Db	3339	GACATTATAAAGTGTAGTACATTTTGTGTACAGATTTGATATTTCCAAATTTGATGGATGG	3398
QY	3539	GAGGAGAGTGTCTTTAAGCTGTAGGCTTTCTTTGTACTGCAATTTTATAGAGATTTAGCT	3598
Db	3399	GAGGAGAGTGTCTTTAAGCTGTAGGCTTTCTTTGTACTGCAATTTTATAGAGATTTAGCT	3458
QY	3599	TTAATATTTTTAGAGATGTAAAAACATTTCTGCTTTCTTAGTCTTACCTAGTCTGAAACAT	3658

Db	3459	TTAATATTTTATAGAGATGAACACATCTGCTTTCTTAGCTTACCTAGTCTGAAACAT	3518
Qy	3659	TTTATTCAATAAGATTTAATTAATAATTG	3690
Db	3519	TTTATTCAATAAGATTTAATTAATAATTG	3550
RESULT 5			
AK000335			
LOCUS Homo sapiens cDNA FLJ20328 fis, clone HEP10039.			
DEFINITION AK000335			
ACCESSION AK000335			
VERSION AK000335.1 GI:7020350			
KEYWORDS oligo capping, fis (full insert sequence).			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (sites)			
Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,			
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,			
Suzuki,Y., Obayashi,M., Nishi,T., Shibahata,T., Tanaka,T.,			
Nakamura,Y., Isogai,T. and Sugano,S.			
NEDO human cDNA sequencing project			
Unpublished			
2 (bases 1 to 3277)			
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,			
Shibahara,T., Tanaka,T. and Nakamura,Y.			
Direct Submission			
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,			
University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,			
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ms.u-tokyo.ac.jp,			
Tel:81-3-5449-5286, Fax:81-3-5449-5416)			
NEDO human cDNA sequencing project supported by Ministry of			
International Trade and Industry of Japan; cDNA full insert			
sequencing; Research Association for Biotechnology; cDNA library			
construction, 5'- & 3'-end one pass sequencing; Department of			
Virology and Human Genome Center, Institute of Medical Science,			
University of Tokyo (partly supported by Science and Technology			
Agency).			
FEATURES			
Location/Qualifiers			
1..3277			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="HEP10039"			
/cell_line="RepG2"			
/cell_type="Hepatoma"			
/clone_lib="HEP"			
/note="cloning vector pME18SFL3"			
179..1978			
/note="unnamed protein product"			
/codon_start=1			
/protein_id="BAA91092.1"			
/db_xref="GI:7020351"			
/translation="MENCLGESRHEVKEKSIENSTASGKIERYVPLNRLKWPPEKG			
EPTQKILRAQSRSASGRKISSENSYLDLEIGPOLSSSTPFSKNESENNLELRL			
SEISIKDRMAKYQAASVKQSSNTYNNELKASGCEIKHMEQENVPFGEVCIHQ			
EGSKISANENSLAVRSTPADSDRSQVSKVEQVHPKPLSPDSRSLSSESPPKA			
MKXFPARTCEQKTVPMERLLANQQVPHISGFRCSYCNKLSLGTYSLHGRI			
YCKPHNQLFKSGNYDEGFGRPHKDLMAKNEEILERPAOLANARETSPGVE			
DPIAKVGVLAAGMEAKASQOEKDKPAETKKLRTAMPPTTELGGSLBEGIKMS			
KPKWPPDEISKEVPEVDPLDLKKLRSSSLKERSRPTVAASFQSTSVKSPKTVSP			
PIKRGWSMSQSESVGGRVAERKQVENAKSKNGNVKTTWNKESKGETGRKSE			
QDVLMEGVYVKEVSEEEQIKRNYDEDEDEE"			
BASE COUNT 1089 a 651 c 703 g 834 t			
ORIGIN			
Query Match 88.0%; Score 3262.2; DB 9; Length 3277;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 3264; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			

Qy	439	AGACCATCTCTGTGTGAAGTCAAGCCAGCCTCTTCTGTGAGCCAAAGCTGACCAAGA	498
Db	1	AGACCATCTCTGTGTGAAGTCAAGCCAGCCTCTTCTGTGAGCCAAAGCTGACCAAGA	60
Qy	499	AGAACAAATCCACCCAGATCTAGACTAGTCACTCTCTGAAGCCCTCGTTCAGGGTCG	558
Db	61	AGAACAAATCCACCCAGATCTAGACTAGTCACTCTCTGAAGCCCTCGTTCAGGGTCG	120
Qy	559	ATATCCCCACATCAAGGACGGTGAAGATCTTAAAGACCACCTCAACAGAAAGTAAAAAAAT	618
Db	121	ATATCCCCACATCAAGGACGGTGAAGATCTTAAAGACCACCTCAACAGAAAGTAAAAAAAT	180
Qy	619	GGAAAAATTGCTTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAACAC	678
Db	181	GGAAAAATTGCTTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAACAC	240
Qy	679	AGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATGT	738
Db	241	AGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATGT	300
Qy	739	TGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGGAAGTCAAGTGG	798
Db	301	TGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGGAAGTCAAGTGG	360
Qy	799	AAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAATAGGCCCCAGGTCAGTT	858
Db	361	AAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAATAGGCCCCAGGTCAGTT	420
Qy	859	GTCATCTTCTACATTGACTCGAGAAAAATAGAGTAGAGCGAAATCTGGAACTTCCACG	918
Db	421	GTCATCTTCTACATTGACTCGAGAAAAATAGAGTAGAGCGAAATCTGGAACTTCCACG	480
Qy	919	CTCTCAGNAACCTCTATAAAGGATCGAATGCCCAAGTACAGGAGCTGTGTCCAAACA	978
Db	481	CTCTCAGNAACCTCTATAAAGGATCGAATGCCCAAGTACAGGAGCTGTGTCCAAACA	540
Qy	979	AAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTC	1038
Db	541	AAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTC	600
Qy	1039	TAAAAATGGAGCAAAAAGGAGAAATGTGCCCCAGGTCCTCAGGTCCTCATCACCCTCAGGA	1098
Db	601	TAAAAATGGAGCAAAAAGGAGAAATGTGCCCCAGGTCCTCAGGTCCTCATCACCCTCAGGA	660
Qy	1099	AGGGAAAAAGATTTCTGCAAAATAGAAATAGCCTGGCAGTCGCTTCCACCCCTGCCGAAGA	1158
Db	661	AGGGAAAAAGATTTCTGCAAAATAGAAATAGCCTGGCAGTCGCTTCCACCCCTGCCGAAGA	720
Qy	1159	TGACTCCCGTGACTCCCGAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACT	1218
Db	721	TGACTCCCGTGACTCCCGAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACT	780
Qy	1219	AAGTCCAGATTCAGAGGCTCCAGTCTTCTGAAAGTTTCTCTCCCAAGCAATGAAGAA	1278
Db	781	AAGTCCAGATTCAGAGGCTCCAGTCTTCTGAAAGTTTCTCTCCCAAGCAATGAAGAA	840
Qy	1279	GTTTCAGGCACTGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTTATCCAAATGGA	1338
Db	841	GTTTCAGGCACTGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTTATCCAAATGGA	900
Qy	1339	GGGTCTTCTGGCCAAACAGCAGGTTGTTTCAATCAGCTGCTTCGGTTCCTCTATTGCAA	1398
Db	901	GGGTCTTCTGGCCAAACAGCAGGTTGTTTCAATCAGCTGCTTCGGTTCCTCTATTGCAA	960
Qy	1399	CAACAAATCAGTCTAGGAAACATATGATCTTTACATGGAAGAATCTATTGTAAGCTCA	1458
Db	961	CAACAAATCAGTCTAGGAAACATATGATCTTTACATGGAAGAATCTATTGTAAGCTCA	1020
Qy	1459	CTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTGGGCACAGACACA	1518
Db	1021	CTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTGGGCACAGACACA	1080

Db 3241 AATTAAATTTGAAAAA 3267

RESULT 6
AX017487 LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 31 from Patent WO9947655.
ACCESSION AX017487
VERSION AX017487.1 GI:10042284
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pillarsky,C.
Human nucleic acid sequences from normal breast tissue
Patent: WO 9947655-A 31 23-SEP-1999.
JOURNAL SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..2783
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 901 a 526 c 609 g 747 t

ORIGIN

Query Match 72.8%; Score 2698.2; DB 6; Length 2783;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1001 AATGAGCTGAAGCCAGTGGTGGCGAATCAAAATTCATAAAATGGAGCAAAAGCAAT 1060
DB 40 AATGAGCTGAAGCCAGTGGTGGCGAATCAAAATTCATAAAATGGAGCAAAAGCAAT 99
QY 1061 GTGCCCCCAGGTCCTGAGGTCTGCATCACCATCAGGAAGGGGAAAAGATTTCGCAAAAT 1120
DB 100 GTGCCCCCAGGTCCTGAGGTCTGCATCACCATCAGGAAGGGGAAAAGATTTCGCAAAAT 159
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 1180
DB 160 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 219
QY 1181 AAGAGTAGGTTTCACAGCCTGTCCATCCCAAGCCTAGTCCAGATTCAGAGCCCTCC 1240
DB 220 AAGAGTAGGTTTCACAGCCTGTCCATCCCAAGCCTAGTCCAGATTCAGAGCCCTCC 279
QY 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 1300
DB 280 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 339
QY 1301 ACCTGCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTTGGCCAAACAGCAG 1360
DB 340 ACCTGCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTTGGCCAAACAGCAG 399
QY 1361 GTGTTTCACATCAGCTGCTCCGTTGCTCTCTATTGGAACAAACAACTCAGTCTAGGAACA 1420
DB 400 GTGTTTCACATCAGCTGCTCCGTTGCTCTCTATTGGAACAAACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAAGAAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
DB 460 TATGCATCTTTACATGGAAGAAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 1540
DB 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 579
QY 1541 AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGAGACCCCT 1600
DB 580 AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGAGACCCCT 639

QY 1601 CACAGCCCAGGGGTAGAAATGCCCTATTGTCTAAGTGGGTGCTCTGCTGCTCAAGTATG 1660
DB 640 CACAGCCCAGGGGTAGAAATGCCCTATTGTCTAAGTGGGTGCTCTGCTGCTCAAGTATG 699
QY 1661 GAAGCCAAAGCCCTCTCTCAGCAGGAGGAAGACAAAGCCAGCTGAAACCAAGAACTG 1720
DB 700 GAAGCCAAAGCCCTCTCTCAGCAGGAGGAAGGAACAAGCCAGCTGAAACCAAGAACTG 759
QY 1721 AGGATCGCTGGCCACCCCTCTGAAGTTTGAAGTTTCAAGGAGTCCCTTTGAGGAAGGG 1780
DB 760 AGGATCGCTGGCCACCCCTCTGAAGTTTGAAGTTTCAAGGAGTCCCTTTGAGGAAGGG 819
QY 1781 ATCAAAATGTCAAAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCT 1840
DB 820 ATCAAAATGTCAAAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCT 879
QY 1841 GAGGATGTGATCTAGATCTGAAGAAGCTTAAGACAGATCTTCTTCACTGAAGGAAGAAGC 1900
DB 880 GAGGATGTGATCTAGATCTGAAGAAGCTTAAGACAGATCTTCTTCACTGAAGGAAGAAGC 939
QY 1901 CGCCCATTCACGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAACTGTG 1960
DB 940 CGCCCATTCACGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAACTGTG 999
QY 1961 TCCCCACCTATCAGGAAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGA 2020
DB 1000 TCCCCACCTATCAGGAAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGA 1059
QY 2021 AGATTGTCAGAAAGGAAACAAGTGGAAAAATGCCAAGGCTTTTAAAGAAATCGGAATGTG 2080
DB 1060 AGATTGTCAGAAAGGAAACAAGTGGAAAAATGCCAAGGCTTTTAAAGAAATCGGAATGTG 1119
QY 2081 GGAAGAAACAACCTGGCAAAACAAGAAATCTAAAGAGACAGAGGAGAGAAAGTAAGGAA 2140
DB 1120 GGAAGAAACAACCTGGCAAAACAAGAAATCTAAAGAGAGACAGAGGAGAGAAAGTAAGGAA 1179
QY 2141 GGTCTAGTTTGGAGATGAGAAATGAGAAATCTGTAGAAAATGGTGACACTCCGATGAA 2200
DB 1180 GGTCTAGTTTGGAGATGAGAAATGAGAAATCTGTAGAAAATGGTGACACTCCGATGAA 1239
QY 2201 GATGATTAACAGCTTCTCTCAACAAACAATCTCCAAGAAACCCCAAGTCTCTGAATGGTGC 2260
DB 1240 GATGATTAACAGCTTCTCTCAACAAACAATCTCCAAGAAACCCCAAGTCTCTGAATGGTGC 1299
QY 2261 AGTTTGTGAGACAAACCTTTGCTGAAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 2320
DB 1300 AGTTTGTGAGACAAACCTTTGCTGAAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 1359
QY 2321 GTGGAACCTCTGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAGAGA 2380
DB 1360 GTGGAACCTCTGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAGAGA 1419
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCTT 2440
DB 1420 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCTT 1479
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCAATGAAGCAG 2500
DB 1480 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCAATGAAGCAG 1539
QY 2501 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 2560
DB 1540 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 1599
QY 2561 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 2620
DB 1600 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 1659
QY 2621 TTTAAATCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2680
DB 1660 TTTAAATCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1719

Qy	2681	AAATATTCCACCTGATTAATAGCCAGAGATCTCTACTGTATTCGCCAAAGGCAATATTAAAGGT	274
Db	1720	AAATATTCCACCTGATTAATAGCCAGAGATCTCTACTGTATTCGCCAAAGGCAATATTAAAGGT	1779
Qy	2741	AGATAGATGATTAGTAGTATATTGTTTACACACATATTTTGGAAATTAGAGAACATACAGAAG	2800
Db	1780	AGATAGATGATTAGTAGTATATTGTTTACACACATATTTTGGAAATTAGAGAACATACAGAAG	1839
Qy	2801	GAATTTAGGGGCTTAAACATTTACGACATGAATGCACCTTTAGTATTAAGGGGCACAGTTTGTA	2860
Db	1840	GAATTTAGGGGCTTAAACATTTACGACATGAATGCACCTTTAGTATTAAGGGGCACAGTTTGTA	1899
Qy	2861	TATTTTAAATAGAAATACCAATTTAAATTTTATTTAGTATTTTACCTGTTAAGAGATTAATTTAGT	2920
Db	1900	TATTTTAAATAGAAATACCAATTTAAATTTTATTTAGTATTTTACCTGTTAAGAGATTAATTTAGT	1959
Qy	2921	CTTTAAATTTTTTAGGTTAAATTTCTTGCCTGTGATATATATAGGAATTTTACTACTTTTAT	2980
Db	1960	CTTTAAATTTTTTAGGTTAAATTTCTTGCCTGTGATATATATAGGAATTTTACTACTTTTAT	2019
Qy	2981	GTCTGCTCTCTAAACATACATCTCTGAACCTCGACGCTCTGAGGTATATAACAAGAGCAC	3040
Db	2020	GTCTGCTCTCTAAACATACATCTCTGAACCTCGACGCTCTGAGGTATATAACAAGAGCAC	2079
Qy	3041	TTTTTGAGGCAATTTGAAACCAACCTACACTCTTCGGTGCTTTAGAGAGATCTGCTGTCT	3100
Db	2080	TTTTTGAGGCAATTTGAAACCAACCTACACTCTTCGGTGCTTTAGAGAGATCTGCTGTCT	2139
Qy	3101	CCCAATTAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTGCTTCTCTTT	3160
Db	2140	CCCAATTAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTGCTTCTCTTT	2199
Qy	3161	TCGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC	3220
Db	2200	TCGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC	2259
Qy	3221	GGGATCACTGTCCCCCATCTTCGGTGTAGAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA	3280
Db	2260	GGGATCACTGTCCCCCATCTTCGGTGTAGAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA	2319
Qy	3281	GAAGAAGACTGTCTTACACCACCTTGAGCTCAGACCTCTAACCCTGTATTTCCCTTATGAT	3340
Db	2320	GAAGAAGACTGTCTTACACCACCTTGAGCTCAGACCTCTAACCCTGTATTTCCCTTATGAT	2379
Qy	3341	GTCCCTTTTTTGAGACACTAAATTTTTAAATACTTACTAGCTCTGAAATATATTGATTTTT	3400
Db	2380	GTCCCTTTTTTGAGACACTAAATTTTTAAATACTTACTAGCTCTGAAATATATTGATTTTT	2439
Qy	3401	ATCACAGTATCTCAGGGTGAATTTAAACCAACTATAGGCCTTTTCTTGGGATGATTTT	3460
Db	2440	ATCACAGTATCTCAGGGTGAATTTAAACCAACTATAGGCCTTTTCTTGGGATGATTTT	2499
Qy	3461	CTAGCTTTAAGGTTTGGGGACATTTAAACTTCAGTACATTTGCTGTACACAGTTGATAT	3520
Db	2500	CTAGCTTTAAGGTTTGGGGACATTTAAACTTCAGTACATTTGCTGTACACAGTTGATAT	2559
Qy	3521	TCCAAATTGTATGGATGGGAGGAGAGGCTGTCTTAAGCTGTAGGCTTTCTTTTGTACTGCG	3580
Db	2560	TCCAAATTGTATGGATGGGAGGAGAGGCTGTCTTAAGCTGTAGGCTTTCTTTTGTACTGCG	2619
Qy	3581	ATTTATAGAGATTTAGCTTTTAATTTTTTTTAGAGATGTAAACCATTTCTGCTTTCTTTAGTC	3640
Db	2620	ATTTATAGAGATTTAGCTTTTAATTTTTTTTAGAGATGTAAACCATTTCTGCTTTCTTTAGTC	2679
Qy	3641	TTACCTAGCTGAAACATTTTTTATTCAAATAAGATTTTTTAATTAATAATTTGAAAAA	3700
Db	2680	TTACCTAGCTGAAACATTTTTTATTCAAATAAGATTTTTTAATTAATAATTTGAAAAA	2739
Qy	3701	AAA 3703	
Db	2740	AAA 2742	

[illegible]

Db 400 GTGTTTCACATCAGCTGCTCCGTTGGTCTCTATTGCAACAACTCAGTCTAGGAACA 459
Qy 1421 TATGCATCTTTACATGGAAGATCTATTGTGAAGCTCACITCAATCAACTCTTTAAATCT 1480
Db 460 TATGCATCTTTACATGGAAGATCTATTGTGAAGCTCACITCAATCAACTCTTTAAATCT 519
Qy 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCGAAGCAAA 1540
Db 520 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCGAAGCAAA 579
Qy 1541 AATGAAACGGAAGATTTTGGAGAGACCAGCCAGCTTGCMAATGCAAGGGAGACCCCT 1600
Db 580 AATGAAACGGAAGATTTTGGAGAGACCAGCCAGCTTGCMAATGCAAGGGAGACCCCT 639
Qy 1601 CACAGCCAGGGGTAGAAGATGCCCCCTATTGCTTAAGTGGGTGCTCGGTGCMAGTATG 1660
Db 640 CACAGCCAGGGGTAGAAGATGCCCCCTATTGCTTAAGGGGGGTGCTCGGTGCAAGTATG 699
Qy 1661 GAAGCCAAAGGCTCCTCTCACAGAGAGGAAGACAAAGCCAGCTGAAACCAAGAACTG 1720
Db 700 GAAGCCAAAGGCTCCTCTCACAGAGAGGAAGACAAAGCCAGCTGAAACCAAGAACTG 759
Qy 1721 AGGATCGCTCGGCCACCCCCCCTGAACTTGGAAAGTTTCAGGAAGTGCCTTGGAGGAAGG 1780
Db 760 AGGATCGCTCGGCCACCCCCCCTGAACTTGGAAAGTTTCAGGAAGTGCCTTGGAGGAAGG 819
Qy 1781 ATCAAAATGTCAAAGCCCAATGCGCTCCTGAAAGCAAGATCAGCAAGCCCGAAGTTCT 1840
Db 820 ATCAAAATGTCAAAGCCCAATGCGCTCCTGAAAGCAAGATCAGCAAGCCCGAAGTTCT 879
Qy 1841 GAGGATGCGATCTAGATCTCAAGAACTAGACGATCTTCTTCACTGAGGAAGGAAGC 1900
Db 880 GAGGATGCGATCTAGATCTCAAGAACTAGACGATCTTCTTCACTGAGGAAGGAAGC 939
Qy 1901 CGCCCAITCACTGTAGCAGCTTCAATTTCAAAGCACCCTCTGTCAAGAGCCCAAAACTGTG 1960
Db 940 CGCCCAITCACTGTAGCAGCTTCAATTTCAAAGCACCCTCTGTCAAGAGCCCAAAACTGTG 999
Qy 1961 TCCCCCACTATCAGGAAGGCTGAGCATGTCTCAGAGAGCTGAAGAGTCTGTGGGTGGA 2020
Db 1000 TCCCCCACTATCAGGAAGGCTGAGCATGTCTCAGAGAGCTGAAGAGTCTGTGGGTGGA 1059
Qy 2021 AGAGTTGAGAAAGGAACAACTGGAATGCCAAGCTTCTTAAGAGAAATGGAATGTG 2080
Db 1060 AGAGTTGAGAAAGGAACAACTGGAATGCCAAGCTTCTTAAGAGAAATGGAATGTG 1119
Qy 2081 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGACAGGGAAGAGAACTAAGGAA 2140
Db 1120 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGACAGGGAAGAGAACTAAGGAA 1179
Qy 2141 GGTCAATGTTTGGAGATGAGAAATGAGAAATCTGTAGAAAAATGGTGCAGACTCCGATGAA 2200
Db 1180 GGTCAATGTTTGGAGATGAGAAATGAGAAATCTGTAGAAAAATGGTGCAGACTCCGATGAA 1239
Qy 2201 GATGATAACAGCTTCTTCAACACAACTCTCCCAAGAACCCCAAGTCTCTGAATTTGTCG 2260
Db 1240 GATGATAACAGCTTCTTCAACACAACTCTCCCAAGAACCCCAAGTCTCTGAATTTGTCG 1299
Qy 2261 AGTTTGTAGACAACACCTTTGCTGAAGAATTTCACTACTCAGATCAGAAATCCAGGAT 2320
Db 1300 AGTTTGTAGACAACACCTTTGCTGAAGAATTTCACTACTCAGAAATCAGAAATCCAGGAT 1359
Qy 2321 GTGGAATCTCGGAGGGAAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2380
Db 1360 GTGGAATCTCGGAGGGAAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 1419
Qy 2381 AATCGGATTTATGATGAGGATGAGGATGAAGAGTGAACAATTTGCAATGATGCTGGGCCCT 2440
Db 1420 AATCGGATTTATGATGAGGATGAGGATGAAGAGTGAACAATTTGCAATGATGCTGGGCCCT 1479
Qy 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTGTGCAAAATGTGATGCACATAAGCAG 2500
Db 1480 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTGTGCAAAATGTGATGCACATAAGCAG 1539

Qy 2501 GTATCCAGCATGAATGTAATTTTACTTGGAGTAACCTTTGGAAAGAACTTCTTCTTAA 2560
Db 1540 GTATCCAGCATGAATGTAATTTTACTTGGAGTAACCTTTGGAAAGAACTTCTTCTTAA 1599
Qy 2561 AATCAAAAAACAAAAACAAAAACAAAAACACATTTCTAAATACTAGAGATAACTTTAC 2620
Db 1600 AATCAAAAAACAAAAACAAAAACAAAAACACATTTCTAAATACTAGAGATAACTTTAC 1659
Qy 2621 TTAATTTCTTCATTTTAGCAGTGATGATGATGAATGCTGTAAAGCTTGTAACTGGG 2680
Db 1660 TTAATTTCTTCATTTTAGCAGTGATGATGATGAATGCTGTAAAGCTTGTAACTGGG 1719
Qy 2681 AATATTTCCACCTGATAATAGCCAGATCTACTGTATTTCCAAAAAGGCAATATTAAAGT 2740
Db 1720 AATATTTCCACCTGATAATAGCCAGATCTACTGTATTTCCAAAAAGGCAATATTAAAGT 1779
Qy 2741 AGATAGATGATTAGTAGTATATTGTTTACACACTATTTTGGAAATTAGAGAACATACAGAAG 2800
Db 1780 AGATAGATGATTAGTAGTATATTGTTTACACACTATTTTGGAAATTAGAGAACATACAGAAG 1839
Qy 2801 GAATTTAGGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGSCACAGTTTGT 2860
Db 1840 GAATTTAGGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGSCACAGTTTGT 1899
Qy 2861 TATTTTAAATGAATACCAATTTAATTTTTRAGTATTTTACCTGTTAAAGAGATATTAGT 2920
Db 1900 TATTTTAAATGAATACCAATTTAATTTTTRAGTATTTTACCTGTTAAAGAGATATTAGT 1959
Qy 2921 CTTTAAATTTTTRAGTATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2980
Db 1960 CTTTAAATTTTTRAGTATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2019
Qy 2981 GTCCTGCTCTTAACTACATCCTGAATCGACGCTCTGAGGTATAATACACAGAGCAC 3040
Db 2020 GTCCTGCTCTTAACTACATCCTGAATCGACGCTCTGAGGTATAATACACAGAGCAC 2079
Qy 3041 TTTTGAAGCAATGAAAAACCAACTACACTCTTCGCTGCTTAGAGAGATCTGCTGCT 3100
Db 2080 TTTTGAAGCAATGAAAAACCAACTACACTCTTCGCTGCTTAGAGAGATCTGCTGCT 2139
Qy 3101 CCCAAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 3160
Db 2140 CCCAAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 2199
Qy 3161 TCTGTGTATCTGCTCTCATTAATTAAGTGAAGCTGCAATATTTTAGTAATACCTTC 3220
Db 2200 TCTGTGTATCTGCTCTCATTAATTAAGTGAAGCTGCAATATTTTAGTAATACCTTC 2259
Qy 3221 GGGATCACCTGTCCTCCCATCTTCCGTTTGAAGTAAAGTGAAGAGTTTAAAGGAGAGAA 3280
Db 2260 GGGATCACCTGTCCTCCCATCTTCCGTTTGAAGTAAAGTGAAGAGTTTAAAGGAGAGAA 2319
Qy 3281 GAAAGAACTGCTTACACCACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 3340
Db 2320 GAAAGAACTGCTTACACCACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 2379
Qy 3341 GTCCCCCTTTTTCAGACACTAATTTTTTAAATCTACTAGCTCTGAAATATATTGATTTT 3400
Db 2380 GTCCCCCTTTTTCAGACACTAATTTTTTAAATCTACTAGCTCTGAAATATATTGATTTT 2439
Qy 3401 ATCAGAGATTTCTCAGGTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATCATTTT 3460
Db 2440 ATCAGAGATTTCTCAGGTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATCATTTT 2499
Qy 3461 CTAGTCTTAAAGTTTGGGACATTTAAACCTTTGAGTACATTTGTTGTACACAGTTGATAT 3520
Db 2500 CTAGTCTTAAAGTTTGGGACATTTAAACCTTTGAGTACATTTGTTGTACACAGTTGATAT 2559
Qy 3521 TCCAAATTTGATGGAATGGAGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCTTTGACTGC 3580
Db 2560 TCCAAATTTGATGGAATGGAGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCTTTGACTGC 2619

QY	2201	GATGATTAACAGGTTCTCTCAAAACAACAATCTCTCCAAGAAGCCCAAGTCTCTGAAATTGGTGC	2260
DB	1849	GATGATTAACAGGTTCTCTCAAAACAACAATCTCTCCAAGAAGCCCAAGTCTCTGAAATTGGTGC	1908
QY	2261	AGTTTTGTAGACAACACCTTTGCTGAGAATTTCACTACTCAGATCAGAAATCCCAGGAT	2320
DB	1909	AGTTTTGTAGACAACACCTTTGCTGAGAATTTCACTACTCAGAAATCAGAAATCCCAGGAT	1968
QY	2321	GTGAAACTCTGGGAGGGAGAAGTGGTCAAGAGGCTCTCTGTGGGAAGCAACAGATAAAGAGA	2380
DB	1969	GTGAAACTCTGGGAGGGAGAAGTGGTCAAGAGGCTCTCTGTGGGAAGCAACAGATAAAGAGA	2028
QY	2381	AATCGGTATTATGATGAGGATGAGAGTGAAGAGTGACAATTTGCAATGATGCTGGGCGCTT	2440
DB	2029	AATCGGTATTATGATGAGGATGAGAGTGAAGAGTGACAATTTGCAATGATGCTGGGCGCTT	2088
QY	2441	AAATTCATGTTAGTGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG	2500
DB	2089	AAATTCATGTTAGTGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG	2148
QY	2501	GTATCCCGAGCTGAATGTAAATTTACTTTGGAAGTAACCTTTGGAAAAAGAAATTCCTCTCTAA	2560
DB	2149	GTATCCCGAGCTGAATGTAAATTTACTTTGGAAGTAACCTTTGGAAAAAGAAATTCCTCTCTAA	2208
QY	2561	AATCAAAAACAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGACATACTTTAC	2620
DB	2209	AATCAAAAACAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGACATACTTTAC	2268
QY	2621	TTAAATCTTCAATTTTAGCAGCTGATGATATGCATAAGTGCTGTGAAGGCTTTGTAACCTGGGG	2680
DB	2269	TTAAATCTTCAATTTTAGCAGCTGATGATATGCATAAGTGCTGTGAAGGCTTTGTAACCTGGGG	2328
QY	2691	AAATATCCACCTGATTAATAGCCGAGATCTPACGTATTTCCAAAAAGCAATATTAAAGT	2740
DB	2329	AAATATCCACCTGATTAATAGCCGAGATCTPACGTATTTCCAAAAAGCAATATTAAAGT	2388
QY	2741	AGATAGATGATTAGTGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAG	2800
DB	2389	AGATAGATGATTAGTGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAG	2448
QY	2801	GAATTTAGGGGCTTAAACATTACGACCTGAATGCACCTTTAGTATATAAGGGGCACAGTTGTA	2860
DB	2449	GAATTTAGGGGCTTAAACATTACGACCTGAATGCACCTTTAGTATATAAGGGGCACAGTTGTA	2508
QY	2861	TATTTTTAAATGAATACCAATTTAAATTTTTTTAGTATTTTACCTGTTAAGAGATATTAGT	2920
DB	2509	TATTTTTAAATGAATACCAATTTAAATTTTTTTAGTATTTTACCTGTTAAGAGATATTAGT	2568
QY	2921	CTTTAAATTTTTTAGGTTAAATTTCTTGCTGTGTATATATAGAGATTTTACTACTTTAT	2980
DB	2569	CTTTAAATTTTTTAGGTTAAATTTCTTGCTGTGTATATATAGAGATTTTACTACTTTAT	2628
QY	2981	GTCCTGCTCTCTAAACTACATCTCTGAACCTCGACGCTCTGAGGTATATAACAACAGAGCAC	3040
DB	2629	GTCCTGCTCTCTAAACTACATCTCTGAACCTCGACGCTCTGAGGTATATAACAACAGAGCAC	2688
QY	3041	TTTTTGGGCAATTTGAAAAACCAACCTTACACTCTTTCCGTGCTTAGAGAGATCTGCTGTCT	3100
DB	2689	TTTTTGGGCAATTTGAAAAACCAACCTTACACTCTTTCCGTGCTTAGAGAGATCTGCTGTCT	2748
QY	3101	CCCAAAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTT	3160
DB	2749	CCCAAAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTT	2808
QY	3161	TCTGGT-GATATCTGCTCTCTCATAAATTTACTGAAAGCTGCAATTTTTAGTAATACCTT	3219
DB	2809	TCTGGTGGATATCTGCTCTCTCATAAATTTACTGAAAGCTGCAATTTTTAGTAATACCTT	2868
QY	3220	CGGGATCACTGTCCCCCATCTTCCGTGTAGACCAAGTGAAAGTTTTAAGGAGGAAGA	3279
DB	2869	CGGGATCACTGTCCCCCATCTTCCGTGTAGACCAAGTGAAAGTTTTAAGGAGGAAGA	2928

Qy	3280	AGAAAGAACTGCTTTACACCACTTTGAGCTCAGACCTCTCTAAACCCCTGATTTCCCTTATGA	3339
Db	2929	AGAAAGAACTGCTTTACACCACTTTGAGCTCAGACCTCTCTAAACCCCTGATTTCCCTTATGA	2988
Qy	3340	TGTCCTCTTTTGAGACACTAAATTTTAAATACCTTACTAGCTCTGAAATATATTTGATTTT	3399
Db	2989	TGTCCTCTTTTGAGACACTAAATTTTAAATACCTTACTAGCTCTGAAATATATTTGATTTT	3048
Qy	3400	TATCACAGTATTTCTCAGGGTGAATTAACCAACTATAGCCCTTTTCTTTGGGATCATTT	3459
Db	3049	TATCACAGTATTTCTCAGGGTGAATTAACCAACTATAGCCCTTTTCTTTGGGATCATTT	3108
Qy	3460	TCTAGCTTTAAGTTTGGGACATTAATACTTGAGTACATTTTGTGTACACAGTTGATA	3519
Db	3109	TCTAGCTTTAAGTTTGGGACATTAATACTTGAGTACATTTTGTGTACACAGTTGATA	3168
Qy	3520	TTCCAAATTCATGGATGGGAGGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTG	3579
Db	3169	TTCCAAATTCATGGATGGGAGGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTG	3228
Qy	3580	CATTATTAGAGATTTAGCTTTAATATTTTTTACAGATGTAAACATTTCTGCTTTCTTAGT	3639
Db	3229	CATTATTAGAGATTTAGCTTTAATATTTTTTACAGATGTAAACATTTCTGCTTTCTTAGT	3288
Qy	3640	CTTACCTAGTCTGAACATTTTTTATCAATAAGATTTTAAATTAATAATTTGAAAAAAA	3699
Db	3289	CTTACCTAGTCTGAACATTTTTTATCAATAAGATTTTAAATTAATAATTTGAAAAAAA	3348
RESULT 9			
AK000057		2667 bp	linear
LOCUS		2667 bp	linear
DEFINITION		Homo sapiens cDNA FLJ20050 fis, clone COL00688.	PRI 22-FEB-2000
ACCESSION		AK000057	
VERSION		AK000057.1	GI:7019894
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (sites)	
		Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.	
TITLE		NEDO human cDNA sequencing project	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2667)	
AUTHORS		Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	
TITLE		Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp).	
JOURNAL		Unpublished	
COMMENT		<p> NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency). </p>	
FEATURES		Location/Qualifiers	
source		1. 2667	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="COL00688"	
		/tissue_type="colon"	
		/clone_lib="COL"	
		/note="cloning vector pME18SFL3"	
CDS		10..1380	
		/note="unnamed protein product"	
		/codon_start=1	

2958	QY	ATATGAGGAATTTTACTACTTTATGTCTCTCTTAAACTACATCCTGAACCTCGACGTC	301
1922	Db	ATATGAGGAATTTTACTACTTTATGTCTCTCTTAAACTACATCCTGAACCTCGACGTC	1981
3018	QY	TGAGGTATAATACAACAGAGACACTTTTTCGAGCAATTTGAAAAACCAACTACACTCTTCG	3077
1982	Db	TGAGGTATAATACAACAGAGACACTTTTTCGAGCAATTTGAAAAACCAACTACACTCTTCG	2041
3078	QY	GTGCTTAGAGAGATCTGCTGCTCCCAAAATAGCTTTTGTATCTGCAGTGAATTTTACTG	3137
2042	Db	GTGCTTAGAGAGATCTGCTGCTCCCAAAATAGCTTTTGTATCTGCAGTGAATTTTACTG	2101
3138	QY	TACTCCAAATGATTGCTTTCTTTCTGCTGATATCTGTCCTTCATATTTACTCTGAAGC	3197
2102	Db	TACTCCAAATGATTGCTTTCTTTCTGCTGATATCTGTCCTTCATATTTACTCTGAAGC	2161
3198	QY	TGCAATATTTTAGTAATACCTTCGGATCA-----CTGTCCTCCCATCTTCCTGTTTGA	3251
2162	Db	TGCAATATTTTAGTAATACCTTCGGATCA-----CTGTCCTCCCATCTTCCTGTTTGA	2217
3252	QY	GCAAGTGAAGAGTTTAAAGGAGAGAGAAGAACTGTCTTTACACCACTTTGAGCTCAG	3311
2218	Db	-----TGAAGAGTTTAAAGGAGAGAGAAGAACTGTCTTTACACCACTTTGAGCTCAG	2271
3312	QY	ACCTCTAAACCCCTGATTTCCCTTTATGATGTCCCTTTTGAGACACTAAATTTTAAATA	3371
2272	Db	ACCTCTAAACCCCTGATTTCCCTTTATGATGTCCCTTTTGAGACACTAAATTTTAAATA	2331
3372	QY	CTTACTAGCTCTGAAATATATTGATTTTATCACAGTATTTCTCAGGGTGAATTTAAACCA	3431
2332	Db	CTTACTAGCTCTGAAATATATTGATTTTATCACAGTATTTCTCAGGGTGAATTTAAACCA	2391
3432	QY	ACTATAGGCTTTTCTTGGATGATTTTCTAGTCTTAAGGTTTGGGACATTTATAAAT	3491
2392	Db	ACTATAGGCTTTTCTTGGATGATTTTCTAGTCTTAAGGTTTGGGACATTTATAAAT	2451
3492	QY	TGAGTACATTTGTTGACAGTTGATATCCAAATTTGATCGATGGCAGGAGAGGTGT	3551
2452	Db	TGAGTACATTTGTTGACAGTTGATATCCAAATTTGATCGATGGCAGGAGAGGTGT	2511
3552	QY	CTTAAGCTGTAGGCTTTCTTTCTAGTCACTTATAGAGATTTAGCTTTAATATTTTTTA	3611
2512	Db	CTTAAGCTGTAGGCTTTCTTTCTAGTCACTTATAGAGATTTAGCTTTAATATTTTTTA	2571
3612	QY	GAGATGTAACAACATCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTCAATAA	3671
2572	Db	GAGATGTAACAACATCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTCAATAA	2631
3672	QY	AGATTTTAATTTAAATTTGAATAAAAAAAAAAAAAA 3705	
2632	Db	AGATTTTAATTTAAATTTGAATAAAAAAAAAAAAAA 2665	
RESULT 10			
AK000372			
LOCUS	AK000372 2536 bp mRNA linear PRI 22-FEB-2000		
DEFINITION	Homo sapiens cDNA FLJ20365 fis, clone HEP17877.		
ACCESSION	AK000372		
VERSION	AK000372.1 GI:7020415		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (sites)		
	Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2536)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T. Isogai,T.,		

Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES
source Location/Qualifiers
1..2536
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP17877"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/note="cloning vector pME189FL3"
85..2199
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91120.1"
/db_xref="GI:7020416"
/translation="MESSPPNRRWTSLSLRVAKELSLVNKNKSSAIVEIFSKYQKAAEETNMKKRNTENLSOHFKGTLTVLKKWENPGLGAESHTSLRNSSTEIRRADHPPEVTSHAASGAKADQEQIHRPSRUPEALVQGRYPHIKQEGDLKDHSTESKRMENCLGESRVEKESISENTDASGKIEKYNVPLNRLXMMFEKGEPTQTKILRAQSRASKRISNSYSLDLEITGPQLASSTFDSEKNESRRNLELPRLSSTGIDKNMAKYQAVSQQSSNTVNTLNASGGGIKHKMEQKNVPPGVCVITHQGEKISANENSLAVRSTPREDSDRSQVSEVQVPHKPLSPDRASSLSSESSPKKMKQAPARETCVQKTVYMERLLAQWQPHISCFRCSPYCNKLSUGTYASHGRITCKPHNQLFKSKGNYDGFGRPHKLDQWKNSENEELERPAQLANARETPHSGVEDAPDAKGVILAAASMPKASQOQEKDQKPAETKKLIRAMPPTLEGSSALGEEKMSPKMPPEDEIISKPEVEDLDLKLRSRSLKERSPTVAAQSTVSPKPTVSPRIKGSWMSSEQSESVGGRVABRKOVENAKSKKXGNVKTQWQNKESKETGTGKRSKEGSHLEMENENLVN GADSDDEDNSFLKHLHKNPSL"
BASE COUNT 863 a 551 c 612 g 510 t
ORIGIN
Query Match 67.9%; Score 2515.4; DB 9; Length 2536;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2530; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 53 AGAGCGCCGGACCTGTACAGGGCTGTGTAGCAGCGCAGAGAAAGCGGCTTTTAGCCA 112
Db |
QY 113 GGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGGCAATGGACC 172
Db |
QY 61 GGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGGCAATGGACC 120
QY 173 TCACATATCATTGAGGGTAAACAGCAAAAGCACTTCTCTGTTCACCAAGAACCAAGTCATCG 232
Db |
QY 121 TCACATATCATTGAGGGTAACAGCCAAAGCACTTCTCTGTTCACCAAGAACCAAGTCATCG 180
QY 233 GCTATTGTGGAATATTTCTTCCAAGTACCAGAAAGCAGCTGAAGAACCAAAATGCGAAG 292
Db |
QY 181 GCTATTGTGGAATATTTCTTCCAAGTACCAGAAAGCAGCTGAAGAACCAAAATGCGAAG 240
QY 293 AAGAGAACTAACCCGAAAAATCTCTCCAGCACTTTAGAAAAGGGACCCCTGACTGTGTTA 352
Db |
QY 241 AAGAGAACTAACCCGAAAAATCTCTCCAGCACTTTAGAAAAGGGACCCCTGACTGTGTTA 300
QY 353 AAGAGAACTGGGAGAACCCAGGGCTGGAGCAGAGTCTCACAGACTCTCTACCGAAC 412
Db |
QY 301 AAGAGAACTGGGAGAACCCAGGGCTGGAGCAGAGTCTCACAGACTCTCTACCGAAC 360
QY 413 AGCAGCTAGATTAGGCACAGAGCAGCACTCTCTCTGCTGAAGTCAGAACGCCAGCT 472

Db 361 AGCAGCACTGAGATTAGGCACAGAGCAGACCATCTCTCTGCTGAGTGCACAGCCAGCT 420
QY 473 GCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCAACCCAGATCTTATAGATCTCAGGTCA 532
Db 421 GCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCAACCCAGATCTTATAGATCTCAGGTCA 480
QY 533 CTTCTCTGAAGCCCTCGTTGAGGTCATATCCACATCAAGGACCGTGAGGATCTTAAA 592
Db 481 CTTCTCTGAAGCCCTCGTTGAGGTCATATCCACATCAAGGACCGTGAGGATCTTAAA 540
QY 593 GACCACCTCAACAGAAAGTAAATAAATGGAAATTTGTCTAGGAGAAATCCAGGATGAAGTA 652
Db 541 GACCACCTCAACAGAAAGTAAATAAATGGAAATTTGTCTAGGAGAAATCCAGGATGAAGTA 600
QY 653 GAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTT 712
Db 601 GAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTT 660
QY 713 CCCTGAAACAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTTAAGATTTCTC 772
Db 661 CCCTGAAACAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTTAAGATTTCTC 720
QY 773 CGGGCCCAAGCCGAAGTGCAAGTGAAGGAAGATCTCTCAAAACAGCTATTCTCTAGAT 832
Db 721 CGGGCCCAAGCCGAAGTGCAAGTGAAGGAAGATCTCTCAAAACAGCTATTCTCTAGAT 780
QY 833 GACCTGGAATAGGCCAGGTGAGTTGTCATCTTTACATTTTGAACTCGGAGAAAATGAG 892
Db 781 GACCTGGAATAGGCCAGGTGAGTTGTCATCTTTACATTTTGAACTCGGAGAAAATGAG 840
QY 893 AGTAGACGAAATCTGAACTTCACGCTCTCAGAACCTCTATAAGGATCGAATGGCC 952
Db 841 AGTAGACGAAATCTGAACTTCACGCTCTCAGAACCTCTATAAGGATCGAATGGCC 900
QY 1013 GCCAGTGGTGGCAAAATCAAAATTCATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGT 1072
Db 961 GCCAGTGGTGGCAAAATCAAAATTCATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGT 1020
QY 1073 CTTGAGTCTGCAATCACCATCAGGAAGGGGAAAGATTCTGCAAAATGAGAAATAGCCTG 1132
Db 1021 CTTGAGTCTGCAATCACCATCAGGAAGGGGAAAGATTCTGCAAAATGAGAAATAGCCTG 1080
QY 1133 GCAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTACTCCCGATTTAAGAGTGAAGTT 1192
Db 1081 GCAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTACTCCCGATTTAAGAGTGAAGTT 1140
QY 1193 CAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCCAGTCTTTCTGAA 1252
Db 1141 CAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCCAGTCTTTCTGAA 1200
QY 1253 AGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAA 1312
Db 1201 AGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAA 1260
QY 1313 TGTCAAGACAGTCTATCCAAATGGAAGGCTCTCTGGCCAAACAGAGGTGTTTCAATC 1372
Db 1261 TGTCAAGACAGTCTATCCAAATGGAAGGCTCTCTGGCCAAACAGAGGTGTTTCAATC 1320
QY 1373 AGCTGCTTCGGTTCCTTATTCACACCAAACTCAGTCTAGGACATATGCAATCTTA 1432
Db 1321 AGCTGCTTCGGTTCCTTATTCACACCAAACTCAGTCTAGGACATATGCAATCTTA 1380
QY 1433 CATGGAAGAAATCTATTGTAAGCCCTCACTCAATCAACTCTTTAAATCTTAAGGGCACTAT 1492
Db 1381 CATGGAAGAAATCTATTGTAAGCCCTCACTCAATCAACTCTTTAAATCTTAAGGGCACTAT 1440
QY 1493 GATGAAGGCTTTGGGCACAGACACCAAGGATCTATGGGCAAGCAAAATGAAACCGAA 1552

1441 GATGAAGGCTTTGGGCACAGACCACACAAAGGATCTATGGGCAAGCAAAATGAAACCGAA 1500
1553 GAGATTTTGGAGAGACACAGCCAGCTTTGCAAAATGCAAGGGAGAGCCCTCAGAGCCAGGG 1612
1501 GAGATTTTGGAGAGACACAGCCAGCTTTGCAAAATGCAAGGGAGAGCCCTCAGAGCCAGGG 1560
1613 GTAGAAGATGCCCTATTGCTTAAGSTGGTGCTCTGGCTGCAAGTATGGAAGCAAGGCC 1672
1561 GTAGAAGATGCCCTATTGCTTAAGSTGGTGCTCTGGCTGCAAGTATGGAAGCAAGGCC 1620
1673 TCCTCTCAGCAGGAGAGGAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGG 1732
1621 TCCTCTCAGCAGGAGAGGAAGAGACAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGG 1680
1733 CCACCCCCCACTGAACCTTTGGAAGTTTCAGGAAGTTCCTTTGAGGAGGGATCAAAATGTCA 1792
1681 CCACCCCCCACTGAACCTTTGGAAGTTTCAGGAAGTTCCTTTGAGGAGGGATCAAAATGTCA 1740
1793 AAGCCCAATGGCCTCTGGAAGACGAATCAGCAAGCCGAGGTTCTTGAGGATGTGAT 1852
1741 AAGCCCAATGGCCTCTGGAAGACGAATCAGCAAGCCGAGGTTCTTGAGGATGTGAT 1800
1853 CTAGATCTGAAGAAAGCTTAAGACGATCTTTCTCACTGAAGGAAAGAGCCGCCCTTCACT 1912
1801 CTAGATCTGAAGAAAGCTTAAGACGATCTTTCTCACTGAAGGAAAGAGCCGCCCTTCACT 1860
1913 GTAGCAGGCTTCATTTTCAAGACACCTCTGTCAAGAGCCCAAAATCTGTGCCCCCTATC 1972
1861 GTAGCAGGCTTCATTTTCAAGACACCTCTGTCAAGAGCCCAAAATCTGTGTGCCCCCTATC 1920
1973 AGGAAAGGCTGGAGCATGTCTCAGAGCAGAGTCAAGAGTCTGTGGTGGAGAGTTCGACAA 2032
1921 AGGAAAGGCTGGAGCATGTCTCAGAGCAGAGTCAAGAGTCTGTGGTGGAGAGTTCGACAA 1980
2033 AGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACC 2092
1981 AGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACC 2040
2093 TGGCAAAACAAAGATCTAAGAGGAGAGACGGAGAGAGTAAAGAGGTCTAGTTTG 2152
2041 TGGCAAAACAAAGATCTAAGAGGAGAGACGGAGAGAGTAAAGAGGTCTAGTTTG 2100
2153 GAGATGGAGAAATGGAATCTTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGC 2212
2101 GAGATGGAGAAATGGAATCTTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGC 2160
2213 TTCTCTCAAAACAAATCTCCAAAGAACCCAAAGTCTCTGAAATGGTCCAGTTTGTAGAC 2272
2161 TTCTCTCAAAACAAAC-ATCTCCAAAGAACCCAAAGTCTCTGAAATGGTCCAGTTTGTAGAC 2219
2273 AACACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGGATGTCGAACTCTGG 2332
2220 AACACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGGATGTCGAACTCTGG 2279
2333 GAGGAGAAATGGTCAAAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATTCGGTATTAT 2392
2280 GAGGAGAAATGGTCAAAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATTCGGTATTAT 2339
2393 GATGAGGATGAGGATGAGAGTGAACAAATTCGAATGATGTGGGCTTTAAATTCATGTTA 2452
2340 GATGAGGATGAGGATGAGAGTGAACAAATTCGAATGATGTGGGCTTTAAATTCATGTTA 2399
2453 GTGTTAGGAGCCACTGCGCTTTGTCAAAATGTGATGCACATAGCAGGTTATCCAGCAT 2512
2400 GTGTTAGGAGCCACTGCGCTTTGTCAAAATGTGATGCACATAGCAGGTTATCCAGCAT 2459
2513 GAAATGTAATTTACTTGGAACTACTTTGGAAAGAAATTCCTCTTAAATCAAAACAA 2572
2460 GAAATGTAATTTACTTGGAACTACTTTGGAAAGAAATTCCTCTTAAATCAAAACAA 2519
2573 AACAAAAAACACAAAA 2589
2520 AAAAAAANAANAANAANA 2536

RESULT 11	AC008147	115345 bp	DNA	linear	PRI 29-MAR-2003
LOCUS	AC008147/c				
DEFINITION	Homo sapiens 12 BAC RP3-405J10 (Roswell Park Cancer Insititute Human BAC library) complete sequence.				
ACCESSION	AC008147				
VERSION	1				
KEYWORDS	HTG.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 115345)				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,D., Chowdhry,I., Christopoulos,C., Cleveland,D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douchwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emeling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunatratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,T., Liu,W., Loulsegod,H., Lozadok,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondei,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mathiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okwuonu,G., Ogunyemi,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prince,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoostari,N., Sison,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,K., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teiford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Zucherlapati,R., Weinstock,O. and Gibbs.R.				
REFERENCE	Submitted (01-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
AUTHORS	Worley K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 115345)				
AUTHORS	Worley K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	4 (bases 1 to 115345)				
AUTHORS	Worley K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	5 (bases 1 to 115345)				
AUTHORS	Worley K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	On Feb 1, 2002 this sequence version replaced gi:15626025. INFORMATION: http://www.bgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu				
COMMENT	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.				
FEATURES	ANNOTATION OF FEATURES: STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3399-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least 2 exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar. SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage. QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html . QUALSTAT-REPORT: Location/Qualifiers 1. 115345 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /Chromosome="12" /Clone="RP3-405J10" complement(2..2047) /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="19465" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluYb"				


```
repeat_region 3668..4037
/rpt_family="LTR24"
repeat_region 4041..4067
/rpt_family="(TA)n"
repeat_region complement(4068..4242)
/rpt_family="AluSg/x"
repeat_region complement(4243..4539)
/rpt_family="AluSg"
repeat_region complement(4541..4836)
/rpt_family="AluY"
repeat_region complement(4837..4890)
/rpt_family="AluSg/x"
repeat_region complement(4891..5005)
/rpt_family="AluJo"
repeat_region 5006..5118
/rpt_family="AluSx"
repeat_region 5119..5425
/rpt_family="AluSx"
repeat_region 5426..5585
/rpt_family="AluSx"
repeat_region complement(5586..5690)
/rpt_family="AluJo"
repeat_region 5931..6233
/rpt_family="AluY"
repeat_region 8032..8320
/rpt_family="AluSg"
repeat_region 8632..8934
/rpt_family="AluSg"
repeat_region 9461..9759
/rpt_family="AluSx"
repeat_region 9920..9947
/rpt_family="(TTTGG)n"
repeat_region complement(9948..10229)
/rpt_family="AluSg"
repeat_region 11773..11943
/rpt_family="LTR47A"
repeat_region complement(114835..15109)
/rpt_family="AluSg"
repeat_region complement(15112..15412)
/rpt_family="AluSg"
repeat_region complement(15467..15779)
/rpt_family="AluSp"
repeat_region 15783..15818
/rpt_family="(T)n"
repeat_region 15932..16230
/rpt_family="AluJo"
repeat_region 16339..16637
/rpt_family="AluJb"
repeat_region 16840..16933
/rpt_family="AluY"
repeat_region 16934..16956
/rpt_family="(CAAA)n"
repeat_region 16962..17249
/rpt_family="AluSg"
repeat_region 17251..17288
/rpt_family="AT-rich"
repeat_region 18079..18112
/rpt_family="(TG)n"

Query Match 51.7%; Score 2284.4; DB 9; Length 115345;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2294; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1383 GTTGCTCTATTGCAACAACTCAGTCTAGGAACATATGATCTTTACATGGAAGAA 1442
|||
Db 3197 GTTATGCTTTTTTCTTTTAAATGACAGICTAGGAACATATGATCTTTACATGGAAGAA 3138
|||

QY 1443 TCTATTGTAAGCTCTACTTCAATCAACTCTTTAAATCTAAGGCAACTATGATGAAGGCT 1502
|||
Db 3137 TCTATTGTAAGCTCTACTTCAATCAACTCTTTAAATCTAAGGCAACTATGATGAAGGCT 3078
|||

QY 1503 TTGGGCACACACACAGGATCTTGGCAAGCAAAAATGAAACGGAAGAGATTTTGG 1562
|||
```

```
Db 3077 TTGGGCACACACACAGGATCTATGGCAAGCAAAAATGAAACGGAAGAGATTTTGG 3018
QY 1563 AGAGACCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGATG 1622
|||
Db 3017 AGAGACCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGATG 2958
|||
QY 1623 CCCCTATTCTAAGGTGGGTGCTCGGTGCAAGTATGCAAGCCAGGCTCTCTCAGC 1682
|||
Db 2957 CCCCTATTCTAAGGTGGGTGCTCGGTGCAAGTATGCAAGCCAGGCTCTCTCAGC 2898
|||
QY 1683 AGGAGAGGAAGACCAAGCCAGCTGAAACCAAGAACTGAGGATCGCTGGCCACCCCA 1742
|||
Db 2897 AGGAGAGGAAGACCAAGCCAGCTGAAACCAAGAACTGAGGATCGCTGGCCACCCCA 2838
|||
QY 1743 CTGAACCTTGAAGTTCAAGAACTGCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 1802
|||
Db 2837 CTGAACCTTGAAGTTCAAGAACTGCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 2778
|||
QY 1803 GGCCTCTTGAAGTTCAAGAACTGCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 1862
|||
Db 2777 GGCCTCTTGAAGTTCAAGAACTGCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 2718
|||
QY 1863 AGAAGCTAAGACGATCTTCTCACTGAAGGAAGGAGCGCCCATTCACCTGAGAGGCT 1922
|||
Db 2717 AGAAGCTAAGACGATCTTCTCACTGAAGGAAGGAGCGCCCATTCACCTGAGAGGCT 2658
|||
QY 1923 CATTTCAAGACACCTCTGTCAAGAGCCCAAAAATGTGTCTCCCACTATCAAGAAAGGCT 1982
|||
Db 2657 CATTTCAAGACACCTCTGTCAAGAGCCCAAAAATGTGTCTCCCACTATCAAGAAAGGCT 2598
|||
QY 1983 GGAGCATGTCAAGCAGAGTGAAGAGTGTGTGGTGGAGAGTTCAGAAAAGGAACAAG 2042
|||
Db 2597 GGAGCATGTCAAGCAGAGTGAAGAGTGTGTGGTGGAGAGTTCAGAAAAGGAACAAG 2538
|||
QY 2043 TGGAAATGCCAAGGCTTCTAAGAAAGTGGAAATGTGGGAAAAACAACCTGGCAAAACA 2102
|||
Db 2537 TGGAAATGCCAAGGCTTCTAAGAAAGTGGAAATGTGGGAAAAACAACCTGGCAAAACA 2478
|||
QY 2103 AAGAATCTAAGGAGAGACAGGGAAGAGAGTGAAGAAAGGTTCATAGTTGGAGATGGAGA 2162
|||
Db 2477 AAGAATCTAAGGAGAGACAGGGAAGAGAGTGAAGAAAGGTTCATAGTTGGAGATGGAGA 2418
|||
QY 2163 ATGAGAATCTTTAGAAAATGGTGCAGACTCCGATGAAGATGATAAACAGCTTCCTCAAC 2222
|||
Db 2417 ATGAGAATCTTTAGAAAATGGTGCAGACTCCGATGAAGATGATAAACAGCTTCCTCAAC 2358
|||
QY 2223 AACAACTCTCAAGAACCCAAAGTCTCTGAATTTGTCAGTTTTGTAGACAAACCTTTG 2282
|||
Db 2357 AACAACTCTCAAGAACCCAAAGTCTCTGAATTTGTCAGTTTTGTAGACAAACCTTTG 2298
|||
QY 2283 CTGAAGAAATTCACCTACTCAGAAATCCAGGATCCAGGATGTGGAACCTGGGAGGAGAG 2342
|||
Db 2297 CTGAAGAAATTCACCTACTCAGAAATCCAGGATGTGGAACCTGGGAGGAGAGAG 2238
|||
QY 2343 TGGTCAAAAGAGCTCTCTGGAAGAACAGATAAAGAGAAATCGGTATTTATCATGAGGATG 2402
|||
Db 2237 TGGTCAAAAGAGCTCTCTGGAAGAACAGATAAAGAGAAATCGGTATTTATCATGAGGATG 2178
|||
QY 2403 AGGATGAAGAGTGAACAAATTCGAATGTGCGGCCCTTAAATTCATGTTAGTTAGCGA 2462
|||
Db 2177 AGGATGAAGAGTGAACAAATTCGAATGTGCGGCCCTTAAATTCATGTTAGTTAGCGA 2118
|||
QY 2463 GCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGAGGATATCCAGCATGAAATGTAAT 2522
|||
Db 2117 GCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGAGGATATCCAGCATGAAATGTAAT 2058
|||
QY 2523 TTACTTGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAATCAAAAACAAAACAAAAA 2582
|||
Db 2057 TTACTTGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAATCAAAAACAAAACAAAAA 1998
|||
QY 2583 CAAAAAACAACATCTTAAATACTAGAGATAACCTTTACTTAAATTTCTTATTTAGCAGT 2642
|||
Db 1997 CAAAAAACAACATCTTAAATACTAGAGATAACCTTTACTTAAATTTCTTATTTAGCAGT 1938
|||
```

QY	2643	GATGATATGCATAAGTGTCTGAAGGCTTGTAACTGGGGAATATTCCACCTGATTAATAGC	2702
Db	1937	GATGATATCGTAAGTGTCTGAAGGCTTGTAACTGGGGAATATTCCACCTGATTAATAGC	1878
QY	2703	CCAGATTCTACTGTATTCCAAAAGGCAATATTAAAGGTAGATAGATGATTAGTAGTATAT	2762
Db	1877	CCAGATTCTACTGTATTCCAAAAGGCAATATTAAAGGTAGATAGATGATTAGTAGTATAT	1818
QY	2763	TGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGCTTAAACATTA	2822
Db	1817	TGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGCTTAAACATTA	1758
QY	2823	CCAGTGAATGCACTTTAGTATTAAGGGCACAGTTTGTATATTTTAAATGAATACCAATT	2882
Db	1757	CCAGTGAATGCACTTTAGTATTAAGGGCACAGTTTGTATATTTTAAATGAATACCAATT	1698
QY	2883	TAATTTTTAGTATTTTACCTGTTTAAGAGATTTATTTAGTCTTTTAAATTTTTTAGGTTAATT	2942
Db	1697	TAATTTTTAGTATTTTACCTGTTTAAGAGATTTATTTAGTCTTTTAAATTTTTTAGGTTAATT	1638
QY	2943	TTCTTGCTGTATATATATGAGGAATTTACTACTTTTATGCTGCTCTCTAAACTACATC	3002
Db	1637	TTCTTGCTGTATATATATGAGGAATTTACTACTTTTATGCTGCTCTCTAAACTACATC	1578
QY	3003	CTGAACCTCAGCTCTGAGGTATATACAAAGACACTTTTGTAGGGCAATTTGAAAAACC	3062
Db	1577	CTGAACCTCAGCTCTGAGGTATATACAAAGACACTTTTGTAGGGCAATTTGAAAAACC	1518
QY	3063	AACTACACTCTTCGGTGCTTAGAGAGATCTGCTGCTCCCAATAAGCTTTTGTATCTG	3122
Db	1517	AACTACACTCTTCGGTGCTTAGAGAGATCTGCTGCTCCCAATAAGCTTTTGTATCTG	1458
QY	3123	CCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCTGGTGATATCTGTGCTTCTC	3182
Db	1457	CCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCTGGTGATATCTGTGCTTCTC	1398
QY	3183	ATAATTACTGAAAGCTGCAATTTTGTAGTAATACCTTCGGGATCACTGTCCCCCATCTTC	3242
Db	1397	ATAATTACTGAAAGCTGCAATTTTGTAGTAATACCTTCGGGATCACTGTCCCCCATCTTC	1338
QY	3243	CTGTGTAGAGCAAGTGAAGGTTTAAAGGAGGAGAGAGAACTGTCTTACACCACT	3302
Db	1337	CTGTGTAGAGCAAGTGAAGGTTTAAAGGAGGAGAGAGAACTGTCTTACACCACT	1278
QY	3303	TGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTATGATGTCCTCTTTTGTAGACACTAAT	3362
Db	1277	TGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTATGATGTCCTCTTTTGTAGACACTAAT	1218
QY	3363	TTTTAAATCTTACTAGCTCTGAAATATATGATTTTATCACAGTATTTCTCAGGGTGAA	3422
Db	1217	TTTTAAATCTTACTAGCTCTGAAATATATGATTTTATCACAGTATTTCTCAGGGTGAA	1158
QY	3423	ATTAACCAACTATAGGCTTTTCTTGGGATGATTTTCTAGCTTTAAGGTTTGGGGACA	3482
Db	1157	ATTAACCAACTATAGGCTTTTCTTGGGATGATTTTCTAGCTTTAAGGTTTGGGGACA	1098
QY	3483	TTATAAATCTTGTAGTACATTTTGTACACAGTGTATTTCCAAATTTGATCGATGGGAGG	3542
Db	1097	TTATAAATCTTGTAGTACATTTTGTACACAGTGTATTTCCAAATTTGATCGATGGGAGG	1038
QY	3543	GAGAGGTGTCTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTATAGAGATTTAGCTTTAA	3602
Db	1037	GAGAGGTGTCTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTATAGAGATTTAGCTTTAA	978
QY	3603	TATTTTTTAGAGATGTAAACATTTCTGCTTTCTTGTAGCTTTACCTAGTCTGAAACATTTT	3662
Db	977	TATTTTTTAGAGATGTAAACATTTCTGCTTTCTTGTAGCTTTACCTAGTCTGAAACATTTT	918
QY	3663	ATTCAATAAAGATTTTAAATTTAAATTTGAA	3692
Db	917	ATTCAATAAAGATTTTAAATTTAAATTTGAA	888

RESULT 12	BC001247	2164 bp	mRNA	linear	PR1 12-JUL-2001
LOCUS	BC001247				
DEFINITION	Homo sapiens, epithelial protein lost in neoplasm beta, clone				
	MGC:4969 IMAGE:3452714, mRNA, complete cds.				
ACCESSION	BC001247				
VERSION	BC001247.1	GI:12654808			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2164)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-DEC-2000) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk				
	Email: cgabbs-remail.nih.gov				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome				
	Center, Stanford University School of Medicine, Stanford, CA 94305				
	Web site: http://www-shgc.stanford.edu				
	Contact: (Dickson, Mark) mcd@paxil.stanford.edu				
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,				
	R. M.				
Clone distribution: MGC clone distribution information can be found					
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov					
Series: IRAX Plate: 4 Row: 0 Column: 9					
This clone was selected for full length sequencing because it					
passed the following selection criteria: matched mRNA gi: 7020350.					
FEATURES	Location/Qualifiers				
source	1..2164				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="LocusID:51474"				
	/db_xref="taxon:9606"				
	/clone="MGC:4969 IMAGE:3452714"				
	/tissue_type="Placenta, choriocarcinoma"				
	/clone_lib="NIH MGC 10"				
	/lab_host="DH10B"				
	/note="Vector: pCMV-SPORT6"				
	187..1983				
	/codon_start=1				
	/product="epithelial protein lost in neoplasm beta"				
	/protein_id="AAH01247.1"				
	/db_xref="GI:12654809"				
	/translation="MENCLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMFPEKG				
	EPQTILKAQSRASGRKISNSYSLSLDELEPGQLSSSTFDEKNEGRNLELRL				
	SETSIDRANKYQAASRQSSSTNYTNELKASGEIKIHMEQKENVPGPEVCITHQ				
	EGKISKANENLAJVRSTPAEDSDRSQVEQVHPKPLSPDSRASLSLSESPKPA				
	MKFQAPARTCEQKTVYPMERLLANQQVHFHISCFCSCYNNKLSLGTYSLSHRIY				
	CKFHNQLFKSGNVDGFGPHKOLWASKNEBELRPAQLANARETPHSPGVED				
	APIAKYGVLAASMEAKASSQEKEDKPAETKRLRIAMPPTTELSSGSALEEGIKMSK				
	KPWPPDELSKEVEPVDVLDLKKRRSSSLKRSRPFVTAAQTSVSKPTKVSPP				
	IRKWSMSQSSVEGVRVAERKVENAKSKNGNKVKTWTWONKESKGETGKRSKEG				
	HSLMENLNLVNGADSDSDFLKQSQPQPKSLNWSFVDNTPAEFTTQNKQSK				
	DVELMEGVVVKLSVBEQIKRNRVYDEDEEE"				
BASE COUNT	744 a	472 c	513 g	435 t	
ORIGIN					
Query Match	57.8%; Score 2141; DB 9; Length 2164;				
Best Local Similarity	99.6%; Pred. No. 0;				
Matches 2157; Conservative	0; Mismatches 5; Indels 3; Gaps 1;				

Db 1 CACAGAGCAGACCATCTCTCTGCTGAAAGTCAAGCCACGCTGCTTCTGAGGCCAAAGCT 60
QY 491 GACCAAGAGAACAAATCCACCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCGTT 550
Db 61 GACCAAGAGAACAAATCCACCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCGTT 120
QY 551 CAGGCTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGABAAGT 610
Db 121 CAGGCTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGABAAGT 180
QY 611 AAAAAAATGGAAAAATGTCTAGGAGATCCAGGCGATGAAGTAGAAAAATCAGAAAAATCAGT 670
Db 181 AAAAAAATGGAAAAATGTCTAGGAGATCCAGGCGATGAAGTAGAAAAATCAGAAAAATCAGT 240
QY 671 GAAAAACACAGATGCTTGGGCAAAATAGAGAAATATATATGTTCCGCTGAAACAGGCTTAAAG 730
Db 241 GAAAAACACAGATGCTTGGGCAAAATAGAGAAATATATATGTTCCGCTGAAACAGGCTTAAAG 300
QY 731 ATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTTCTCGGGGCCCAAGCCGAAAGT 790
Db 301 ATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTTCTCGGGGCCCAAGCCGAAAGT 360
QY 791 GCAAGTGGAAAGGAGATCTCTGAAACACAGCTATTCTCTAGATGACCTTGGAAATAGGCCCA 850
Db 361 GCAAGTGGAAAGGAGATCTCTGAAACACAGCTATTCTCTAGATGACCTTGGAAATAGGCCCA 420
QY 851 GGTCAAGTTGATCTTCTACATTTGATCTCGGAGAAAAATGAGAGTAGACGAAATCTGGAA 910
Db 421 GGTCAAGTTGATCTTCTACATTTGATCTCGGAGAAAAATGAGAGTAGACGAAATCTGGAA 480
QY 911 CTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGCCAAGTACAGGACGCTGTG 970
Db 481 CTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGCCAAGTACAGGACGCTGTG 540
QY 971 TCCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGCGGAATC 1030
Db 541 TCCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGCGGAATC 600
QY 1031 AAAATTCAAAAATGGAGCAAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTGCATCAC 1090
Db 601 AAAATTCAAAAATGGAGCAAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTGCATCAC 660
QY 1091 CATCAGGAAGGGGAAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 1150
Db 661 CATCAGGAAGGGGAAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 720
QY 1151 GCCGAAGATCACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCC 1210
Db 721 GCCGAAGATCACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCC 780
QY 1211 AAGCCACTAAGTCCAGATCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAAGCA 1270
Db 781 AAGCCACTAAGTCCAGATCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAAGCA 840
QY 1271 ATGAAGAAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTGAGAAAGACAGTCTAT 1330
Db 841 AT---GAAAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTGAGAAAGACAGTCTAT 897
QY 1331 CCAATGGAGGCTCTGTGGCCACCGAGGCTTTTCAATCAGTGTCTTCGTTGCTGCC 1390
Db 898 CCAATGGAGGCTCTGTGGCCACCGAGGCTTTTCAATCAGTGTCTTCGTTGCTGCC 957
QY 1391 TATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAGAGATCTATTGT 1450
Db 958 TATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAGAGATCTATTGT 1017
QY 1451 AAGCCTCACTTCAATCAATCTTTTAAATCTAAGGCACTATGATGAAGGCTTTTGGGCAC 1510
Db 1018 AAGCCTCACTTCAATCAATCTTTTAAATCTAAGGCACTATGATGAAGGCTTTTGGGCAC 1077
QY 1511 AGACCACACAAAGATCTATGGGCAAGCAAAAATGAAACGAAAGATTTTGGAGAGACCA 1570

Db 1078 AGACCACACAAAGGATCTATGGGCAAGCAAAAATGAAACGAAAGATTTTGGAGAGACCA 1137
QY 1571 GCCAGCTTTGCAAAATCAAGGGAGAGCCCTCACAGCCAGGGGTAGAAGATGCCCTATT 1630
Db 1138 GCCAGCTTTGCAAAATCAAGGGAGAGCCCTCACAGCCAGGGGTAGAAGATGCCCTATT 1197
QY 1631 GCTAAAGTGGGTGCTCTGCTGCTCAAGATGGAAGCAAGGCTCTCTCAGCAGGAGAG 1690
Db 1198 GCTAAAGTGGGTGCTCTGCTGCTCAAGATGGAAGCAAGGCTCTCTCAGCAGGAGAG 1257
QY 1691 GAAGACAAAGCCAGCTGAAAACCAAGAGCTGAGGATCGCTGGCCACCCCTCACTGAAC 1750
Db 1258 GAAGACAAAGCCAGCTGAAAACCAAGAGCTGAGGATCGCTGGCCACCCCTCACTGAAC 1317
QY 1751 GGAAGTTTCAGGAAGTGCCTTTGGAGGAAGGATCAAAAATGTCAAAGCCCAATGGCTCT 1810
Db 1318 GGAAGTTTCAGGAAGTGCCTTTGGAGGAAGGATCAAAAATGTCAAAGCCCAATGGCTCT 1377
QY 1811 GAAGACGAAATCAGCAAGCCGAAAGTTCCTGAGGATGTCGATCTAGATCTGAAGAAGCTA 1437
Db 1378 GAAGACGAAATCAGCAAGCCGAAAGTTCCTGAGGATGTCGATCTAGATCTGAAGAAGCTA 1437
QY 1871 AGACGATCTTCTCACTGAAGGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAA 1930
Db 1438 AGACGATCTTCTCACTGAAGGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAA 1497
QY 1931 AGACCTCTGTCAAGAGCCCAAAAATCTGTGTCCCACTATCAGCAAAAGGCTGGAGCATG 1990
Db 1498 AGACCTCTGTCAAGAGCCCAAAAATCTGTGTCCCACTATCAGGAAAGGCTGGAGCATG 1557
QY 1991 TCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAACTGCAAAAT 2050
Db 1558 TCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAACTGCAAAAT 1617
QY 2051 GCCAAGGCTTTCAAGAGAAATGGGAATGTGGGAAAAAACAACCTGCAAAACAAAGAAATCT 2110
Db 1618 GCCAAGGCTTTCAAGAGAAATGGGAATGTGGGAAAAAACAACCTGCAAAACAAAGAAATCT 1677
QY 2111 AAAGGAGAGACAGGGAAGAGAGTAAGGAAGTCTAGTTTGGAGATGGAGATCAGAAAT 2170
Db 1678 AAAGGAGAGACAGGGAAGAGAGTAAGGAAGTCTAGTTTGGAGATGGAGATCAGAAAT 1737
QY 2171 CTTGTAGAAAATGGTGCAGACTCCGATCAAGATGATAACAGCTTCTCTCAAAACAACTCT 2230
Db 1738 CTTGTAGAAAATGGTGCAGACTCCGATCAAGATGATAACAGCTTCTCTCAAAACAACTCT 1797
QY 2231 CCACAAAGCCCAAGTCTCTGAAATGGTGCAGTTTGTAGACAACACCTTTGCTCAAGAA 2290
Db 1798 CCACAAAGCCCAAGTCTCTGAAATGGTGCAGTTTGTAGACAACACCTTTGCTCAAGAA 1857
QY 2291 TTCCTACTCAGATCAAGAAATCCAGGATGTGAACTCTCTGGGAGGAGAGTGGTCAAA 2350
Db 1858 TTCCTACTCAGATCAAGAAATCCAGGATGTGAACTCTCTGGGAGGAGAGTGGTCAAA 1917
QY 2351 CAGCTCTCTGTGGAGAACACAGATAAGAGAAATCGGTATTATGATGAGGATCAGAGTAA 2410
Db 1918 CAGCTCTCTGTGGAGAACACAGATAAGAGAAATCGGTATTATGATGAGGATCAGAGTAA 1977
QY 2411 GAGTGACAAAATGCAATGATCTGGGCTTAAATTCATGTTAGTGTAGCGAGGCACTGC 2470
Db 1978 GAGTGACAAAATGCAATGATCTGGGCTTAAATTCATGTTAGTGTAGCGAGGCACTGC 2037
QY 2471 CCTTTGCAAAAATGTGATGACATGAAGCAGGTATCCAGCATGAAATGTAATTTACTTGG 2530
Db 2038 CCTTTGCAAAAATGTGATGACATGAAGCAGGTATCCAGCATGAAATGTAATTTACTTGG 2097
QY 2531 AAGTAACTTTGGAAGAAATCTCTTCTTAAATCAAAAAACAAACAAAAACAAAAA 2590
Db 2098 AAGTAACTTTGGAAGAAATCTCTTCTTAAATCAAAAAACAAACAAAAACAAAAA 2157
QY 2591 ACACA 2595
Db 2158 AAAAA 2162

RESULT 13	BC010664	2121 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS	BC010664				
DEFINITION	Homo sapiens, clone IMAGE:3854371, mRNA, partial cds.				
ACCESSION	BC010664				
VERSION	BC010664.1	GI:14715008			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2121)				
AUTHORS	Straussberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalob@bcm.tmc.edu Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.				
FEATURES	Location/Qualifiers				
source	1..2121				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3854371"				
	/tissue_type="Colon, adenocarcinoma"				
	/clone_lib="NIH_MGC_65"				
	/lab_host="DH10B"				
	/note="Vector: pCMV-SPORT6"				
CDS	<1..831				
	/codon_start=1				
	/product="unknown (protein for IMAGE:3854371)"				
	/protein_id="AAH10664.1"				
	/db_xref="GI:14715009"				
	/translation="ARETPHSPGVDEAPIAKVGVLAASMEAKSSQOEKEDKPAEMKK LRWAPPTEGSGSALDEEGIKMSKPKWPEDEISXPEYEDVDLKLKRSLSLK ERSRPFTVAASFQSTSVKSPVPPRKQMSMEQSEESVGGVRAERKQVENAKASK KXNGVGTWQKSGKEIGKRGKSHLENENLVENGADDEDENSLKQSQPE PXSLNWSFVNTFAEFPTQNKSQDVELMEGEVWKELSVEEIKRNYRVEDDEDE "				
BASE COUNT	697 a	376 c	448 g	600 t	
ORIGIN					
Query Match	57.1%; Score 2115.2; DB 9; Length 2121;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 2117; Conservative	0; Mismatches 3; Indels 0; Gaps 0;				
Qy	1586 GCAAGGAGACCCCTCACAGCCAGGGGTACAGATGCCCTATTGCTAAGGTGGGTGTC 1645				
Db	1 GCAAGGAGACCCCTCACAGCCAGGGGTACAGATGCCCTATTGCTAAGGTGGGTGTC 60				

Qy	1646	CTGGCTGCAAGTATGGAGCCAGGCTCTCTCTCAGCAGGAGAAAGGAGCAAGCCAGCT	1705
Db	61	CTGGCTGCAAGTATGGAGCCAGGCTCTCTCTCAGCAGGAGAAAGGAGCAAGCCAGCT	120
Qy	1706	GAACCCAAAGAGCTGAGATCGCCTGGCCACCCCCCACTGAACCTTGAAGTTCAAGAGT	1765
Db	121	GAACCCAAAGAGCTGAGATCGCCTGGCCACCCCCCACTGAACCTTGAAGTTCAAGAGT	180
Qy	1766	GCCTTGGAGGAGGATCAAAATGTCAGAGCCCAATGGCTCTCTGAGAGCAATCAGC	1825
Db	181	GCCTTGGAGGAGGATCAAAATGTCAGAGCCCAATGGCTCTCTGAGAGCAATCAGC	240
Qy	1826	AAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCA	1885
Db	241	AAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCA	300
Qy	1886	CTGAAGGAAAGACCCGCCCACTTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAG	1945
Db	301	CTGAAGGAAAGACCCGCCCACTTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAG	360
Qy	1946	AGCCCAAAACTGTGTCGCCCACTATCAGGAAAGGCTGGAGCATGTCAGAGCAGAGTGAA	2005
Db	361	AGCCCAAAACTGTGTCGCCCACTATCAGGAAAGGCTGGAGCATGTCAGAGCAGAGTGAA	420
Qy	2006	GAGTCTCTGGGTGGAAGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAG	2065
Db	421	GAGTCTCTGGGTGGAAGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAG	480
Qy	2066	AAGAATGGGAATGTGGGAAAAACAACCTGGCAAAAACAAGAAATCTAAAGGAGAGACAGGG	2125
Db	481	AAGAATGGGAATGTGGGAAAAACAACCTGGCAAAAACAAGAAATCTAAAGGAGAGACAGGG	540
Qy	2126	AAGAGAAGTAAAGGAAGTCTATAGTTTGGAGATGGAGAATGAGAATCTTGTAGAAAATGGT	2185
Db	541	AAGAGAAGTAAAGGAAGTCTATAGTTTGGAGATGGAGAATGAGAATCTTGTAGAAAATGGT	600
Qy	2186	GCAGACTCCGATCAAGATGATACAGCTTCTTCAACACAACTCTCCACAAAGACCCAAAG	2245
Db	601	GCAGACTCCGATCAAGATGATACAGCTTCTTCAACACAACTCTCCACAAAGACCCAAAG	660
Qy	2246	TCTCTGAATTTGGTCGAGTTTGTAGACACACCTTTGCTGAAGAAATCTACTACTCAGAAT	2305
Db	661	TCTCTGAATTTGGTCGAGTTTGTAGACACACCTTTGCTGAAGAAATCTACTACTCAGAAT	720
Qy	2306	CAGAAATCCAGGATGTGGAATCTCTGGAGGGAGAAAGTGGTCAAGAGCTCTCTGTGGAA	2365
Db	721	CAGAAATCCAGGATGTGGAATCTCTGGAGGGAGAAAGTGGTCAAGAGCTCTCTGTGGAA	780
Qy	2366	GAACAGATAAAGAAATCGGTATTATGATGAGGATGAGAGTGAAGAGTGACAAATTGCA	2425
Db	781	GAACAGATAAAGAAATCGGTATTATGATGAGGATGAGAGTGAAGAGTGACAAATTGCA	840
Qy	2426	ATGATGCTGGGCTTAAATTCATTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGT	2485
Db	841	ATGATGCTGGGCTTAAATTCATTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGT	900
Qy	2486	GATGCACATAAGCAGGTATCCCAGCATGAAATGTAATTTACTTGAAGTAACCTTTGGAAA	2545
Db	901	GATGCACATAAGCAGGTATCCCAGCATGAAATGTAATTTACTTGAAGTAACCTTTGGAAA	960
Qy	2546	AGAAATTCCTTTTAAATCAAAAAACAAAAACAAAAACAAAAACAAATCTCTAAATAC	2605
Db	961	AGAAATTCCTTTTAAATCAAAAAACAAAAACAAAAACAAAAACAAATCTCTAAATAC	1020
Qy	2606	TAGAGATACCTTTACTTAAATTTCTTATTTAGCAGTATGATATGCAATAGTGTGTAA	2665
Db	1021	TAGAGATACCTTTACTTAAATTTCTTATTTAGCAGTATGATATGCAATAGTGTGTAA	1080
Qy	2666	GGCTTGTAACTGGGAAATATTCCACCTCGATAATAGCCCAAGTCTACTGTATTTCCAAA	2725
Db	1081	GGCTTGTAACTGGGAAATATTCCACCTCGATAATAGCCCAAGTCTACTGTATTTCCAAA	1140
Qy	2726	AGGCAATTAAGGTAGATAGATGATTAGTAGTATTATTGTACACACTATTTTGGAAATTA	2785

Db	1141	AGGCAATATTAGGTAGATAGATAGTATATTGTTACACACTATTTTGGAAATTA	1200
QY	2786	GAGAACATACAGAGAAATTTAGGGGCTTAAACATTTAGCACTGAATGCACTTTAGTATAA	2845
Db	1201	GAGAACATACAGAGAAATTTAGGGGCTTAAACATTTAGCACTGAATGCACTTTAGTATAA	1260
QY	2846	AGGCACAGTTGTATATTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTT	2905
Db	1261	AGGCACAGTTGTATATTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTT	1320
QY	2906	AAGAGATTATTAGTCTTTTAAATTTTGTAGTATTAAATTTTGTAGTATTACCTGTT	2965
Db	1321	AAGAGATTATTAGTCTTTTAAATTTTGTAGTATTAAATTTTGTAGTATTACCTGTT	1380
QY	2966	AATTTACTACTTTATGTCCTCTCTAACTACATCTGAGCTCGAGCTCTGAGGTAT	3025
Db	1381	AATTTACTACTTTATGTCCTCTCTCTAACTACATCTGAGCTCGAGCTCTGAGGTAT	1440
QY	3026	AATACAACAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGGTGCTTAG	3085
Db	1441	AATACAACAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGGTGCTTAG	1500
QY	3086	AGAGATCTGCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAA	3145
Db	1501	AGAGATCTGCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAA	1560
QY	3146	ATGATTGCTTTCTTCTGCTGATATCTGTGCTTCTCATATTAATTAAGTGAAGAGT	3205
Db	1561	ATGATTGCTTTCTTCTGCTGATATCTGTGCTTCTCATATTAATTAAGTGAAGAGT	1620
QY	3206	TTTAGTAATACCTTCGGGATCACTGTGCCCATCTTCGGTGTAGAGCAAGTGAAGAGT	3265
Db	1621	TTTAGTAATACCTTCGGGATCACTGTGCCCATCTTCGGTGTAGAGCAAGTGAAGAGT	1680
QY	3266	TTAAGGAGGAGAAAGAACTGCTTTACACCACTTGAGCTGAGCTGAGCTGAGCTGAG	3325
Db	1681	TTAAGGAGGAGAAAGAACTGCTTTACACCACTTGAGCTGAGCTGAGCTGAGCTGAG	1740
QY	3326	TATTTCCCTTATGATGTCCTTCTTGGAGCACTAAATTTTAAATCTTACTAGCTCTGA	3385
Db	1741	TATTTCCCTTATGATGTCCTTCTTGGAGCACTAAATTTTAAATCTTACTAGCTCTGA	1800
QY	3386	AATATTGATTTTATCACAGTATCTCAGGCTGAAATTTAAACCACTATAGGCTTTT	3445
Db	1801	AATATTGATTTTATCACAGTATCTCAGGCTGAAATTTAAACCACTATAGGCTTTT	1860
QY	3446	TCCTGGGATGATTTTCTAGTCTTAAGGTTTGGGACATTTAAACCTTGAGTACATTTGTT	3505
Db	1861	TCCTGGGATGATTTTCTAGTCTTAAGGTTTGGGACATTTAAACCTTGAGTACATTTGTT	1920
QY	3506	GTACACAGTTGATATCCAAATTTGATGGATGGAGGGAGAGGTGCTTAAGCTGTAGGC	3565
Db	1921	GTACACAGTTGATATCCAAATTTGATGGATGGAGGGAGAGGTGCTTAAGCTGTAGGC	1980
QY	3566	TTTTCTTTGACTGATTTTATAGAGATTTAGCTTTTAAATTTTATAGAGATGTAAACAT	3625
Db	1981	TTTTCTTTGACTGATTTTATAGAGATTTAGCTTTTAAATTTTATAGAGATGTAAACAT	2040
QY	3626	TCGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTTCAATAAAGATTTTAAATAA	3685
Db	2041	TCGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTTCAATAAAGATTTTAAATAA	2100
QY	3686	ATTGAAAAAATAAAAAA	3705
Db	2101	ATTGAAAAAATAAAAAA	2120

RESULT 14

AF218025

LOCUS

DEFINITION

Homo sapiens clone PP624 unknown mRNA.

ACCESSION

AF218025

PRI 01-OCT-2000

linear

1754 bp

mRNA

PP624

unknown

mRNA.

```
QY 1248 CTGAAGTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGACCTGCG 1307
Db 421 CTGAAGTCTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGACCTGCG 480
QY 1308 TGGAAATGTCAGAGACAGTCTATCCAAATGAGGCTCTCTTGGCCAAACCCAGAGTCTTC 1367
Db 481 TGGAAATGTCAGAGACAGTCTATRCAAATGGAGGCTCTCTTGGCCAAACCCAGAGTCTTC 540
QY 1368 ACATCAGCTGCTCTCCGTTGCTCTATTGCAACCAACCACTCAGTCTAGNACATATGCA 1427
Db 541 ACATCAGCTGCTCTCCGTTGCTCTATTGCAACCAACCACTCAGTCTAGNACATATGCA 600
QY 1428 CTTTACATGAAGAATCTATTGTAAGCTTCACTTCAATCAACTCTTTAAATCTAAGGCCA 1487
Db 601 CTTTACATGAAGAATCTATTGTAAGCTTCACTTCAATCAACTCTTTAAATCTAAGGCCA 660
QY 1488 ACTATGATGAAGGCTTGGGCACAGACCACCAAGGATCTATGGCAACCAAAATGAAA 1547
Db 661 ACTATGATGAAGGCTTGGGCACAGACCACCAAGGATCTATGGCAACCAAAATGAAA 720
QY 1548 ACGAAGAGATTTTGGAGAGACCGCCAGCTTGCAAATGCAAGGGAGACCCCTCAGACC 1607
Db 721 ACGAAGAGATTTTGGAGAGACCGCCAGCTTGCAAATGCAAGGGAGACCCCTCAGACC 780
QY 1608 CAGGGTGAAGATGCCCTATTGCTAAGGTGGTGTCTCTGGCTGCAAGTATGGAAGCCA 1667
Db 781 CAGGGTGAAGATGCCCTATTGCTAAGGTGGTGTCTCTGGCTGCAAGTATGGAAGCCA 840
QY 1668 AGGCTCTCTCAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCG 1727
Db 841 AGGCTCTCTCAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCG 900
QY 1728 CTTGGCCACCCCTGACTTGGAGTTCAGGAGTGCCTTGGAGGAGGATCAAAA 1787
Db 901 CTTGGCCACCCCTGACTTGGAGTTCAGGAGTGCCTTGGAGGAGGATCAAAA 960
QY 1788 TGTCAAAGCCAAATGGCTCTCTGAAGAGCAATCAGCAAGCCGGAAGTCTCTGAGGATG 1847
Db 961 TGTCAAAGCCAAATGGCTCTCTGAAGAGCAATCAGCAAGCCGGAAGTCTCTGAGGATG 1020
QY 1848 TCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCAT 1907
Db 1021 TCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCAT 1080
QY 1908 TCACCTGTACAGCTTCATTTCAAGCACCCTCTGTCAAGAGCCCAAAACTGTGTCCCAAC 1967
Db 1081 TCACCTGTACAGCTTCATTTCAAGCACCCTCTGTCAAGAGCCCAAAACTGTGTCCCAAC 1140
QY 1968 CTATCAGGAAAGGCTGGAGCATGTCTCAGACGAGTGAAGAGTCTGTGGTGGCAAGTTG 2027
Db 1141 CTATCAGGAAAGGCTGGAGCATGTCTCAGACGAGTGAAGAGTCTGTGGTGGCAAGTTG 1200
QY 2028 CAGAAAGGAAAAGTGGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTGGGAAAAA 2087
Db 1201 CAGAAAGGAAAAGTGGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTGGGAAAAA 1260
QY 2088 CAACCTGGCAAAACCAAGAAATCTTAAGGAGACAGGGAAGAGTGAAGAGGTGATA 2147
Db 1261 CAACCTGGCAAAACCAAGAAATCTTAAGGAGAG-CAGGGAAGAGAGTGAAGAGGTGATA 1319
QY 2148 GTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGACACTCCCATGAAGTATA 2207
Db 1320 GTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGACACTCCCATGAAGTATA 1379
QY 2208 ACAGCTTCTCAAAACAACATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTG 2267
Db 1380 ACAGCTTCTCAAAACAACATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTG 1439
QY 2268 TAGAACAACCTTTCCTGAAGAAATCTACTACTCAGAAATCCCAAGATGTGGAAC 2327
Db 1440 TAGAACAACCTTTCCTGAAGAAATCTACTACTCAGAAATCCCAAGATGTGGAAC 1499
QY 2328 TCTGGAGGGAGAGTGTCAAAGAGCTCTCTGTGGAAGAACAGATATAAGAGAAATCGGT 2387
```

```
Db 1500 TCTGGAGGGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATATAAGAGAAATCGGT 1559
QY 2388 ATTATGATGAGGATGAGGATGAGGATGACAAATTTGCAATGATGCTGGCCTTAAATTTCA 2447
Db 1560 ATTATGATGAGGATGAGGATGAGGATGACAAATTTGCAATGATGCTGGCCTTAAATTTCA 1619
QY 2448 TGTATGATGAGGATGAGGATGAGGATGACAAATTTGCAATGATGATGACATGAGGATATCCC 2507
Db 1620 TGTATGATGAGGATGAGGATGAGGATGACAAATTTGCAATGATGATGACATGAGGATATCCC 1679
QY 2508 AGCATGAATGTAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAA 2567
Db 1680 AGCATGAATGTAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAA 1739
QY 2568 AACAAAAACAAAAA 2582
Db 1740 AAAAAA 1754

RESULT 15
BD160117
LOCUS BD160117
DEFINITION BD160117
ACCESSION BD160117
VERSION BD160117.1
KEYWORDS GI:27865875
SOURCE JP 2002191363-A/14960.
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
TITLE Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
JOURNAL Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
COMMENT Primer for synthesizing full-length cDNA and use thereof
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/14960
PD 09-JUL-2002
PF 28-JUL-2000
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TEISUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00
Primer for synthesizing full-length cDNA and use thereof
FT CDS Location/Qualifiers
source
1..2207
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 592 a 438 c 512 g 565 t
ORIGIN
Query Match 42.3%; Score 1568.4; DB 6; Length 2207;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 1563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAGAT 1060
Db 638 AATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAGAT 697
QY 1061 GTCCCCCAGGCTCTGAGGTCTGCATCAACCATCAGGAAGGGGAAAGATTTCTGCAAT 1120
Db 698 GTCCCCCAGGCTCTGAGGTCTGCATCAACCATCAGGAAGGGGAAAGATTTCTGCAAT 757
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGGAGATGACTCCCGTACTCCCGAGTT 1180
```

Db	758	GAGAAATAGCCTGGAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTGACTCCAGGTT	817
Qy	1181	AAGAGTGAAGTTCAACAGAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCC	1240
Db	818	AAGAGTGAAGTTCAACAGAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCC	877
Qy	1241	AGTCTTTCTGAAAGTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACTCTCAAGAGAG	1300
Db	878	AGTCTTTCTGAAAGTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACTCTCAAGAGAG	937
Qy	1301	ACCTGGTGAAGTTCAGAGAGCAGTCTATCCCAATGGAGCGTCTCTTGGCCCAACAGCAG	1360
Db	938	ACCTGGTGAAGTTCAGAGAGCAGTCTATCCCAATGGAGCGTCTCTTGGCCCAACAGCAG	997
Qy	1361	GTGTTTACATCAGCTGCTTCCTGCTGCTATTGCAACAACAACCTCAGTCTAGGAACA	1420
Db	998	GTGTTTACATCAGCTGCTTCCTGCTGCTATTGCAACAACAACCTCAGTCTAGGAACA	1057
Qy	1421	TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACTTCAATCAACTCTTTAAATCT	1480
Db	1058	TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACTTCAATCAACTCTTTAAATCT	1117
Qy	1481	AAGGGCAACTATGATGAAGCCTTTGGGCACAGACACACCAAGGATCTATGGCAAGCAAA	1540
Db	1118	AAGGGCAACTATGATGAAGCCTTTGGGCACAGACACACCAAGGATCTATGGCAAGCAAA	1177
Qy	1541	AATGAAACCAAGAGATTTTGAGAGACAGCCCGAGCTGCAAAATGCAAGGAGACCCCT	1600
Db	1178	AATGAAACCAAGAGATTTTGAGAGACAGCCCGAGCTGCAAAATGCAAGGAGACCCCT	1237
Qy	1601	CACAGCCAGGGGTAGAAGATGCCCTATTGTAAAGTGGGTGCTCTGGCTGCAAGATAG	1660
Db	1238	CACAGCCAGGGGTAGAAGATGCCCTATTGTAAAGTGGGTGCTCTGGCTGCAAGATAG	1297
Qy	1661	GAAGCCAGGCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGCTG	1720
Db	1298	GAAGCCAGGCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGCTG	1357
Qy	1721	AGGATGGCTGCCCCCCTCAACTTGGAACTTGAAGTTCAGGAAGTGCCTTGGAGGAAGG	1780
Db	1358	AGGATGGCTGCCCCCCTCAACTTGGAACTTGAAGTTCAGGAAGTGCCTTGGAGGAAGG	1417
Qy	1781	ATCAAAATGTCAAAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCT	1840
Db	1418	ATCAAAATGTCAAAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCT	1477
Qy	1841	GAGGATGTCGATCTAGATCTGAAGAAAGCTAAGATGATCTTCTTCACTGAAGCAAGAAAGC	1900
Db	1478	GAGGATGTCGATCTAGATCTGAAGAAAGCTAAGATGATCTTCTTCACTGAAGCAAGAAAGC	1537
Qy	1901	CGCCCATTCAGTACAGCTTCTTTCAAGCAGCTCTGTCAAGAGCCCAAAACTGTG	1960
Db	1538	CGCCCATTCAGTACAGCTTCTTTCAAGCAGCTCTGTCAAGAGCCCAAAACTGTG	1597
Qy	1961	TCCCCACCTATCAGGAAAGGTGGAGATGTGAGAGAGTGAAGAGTCTGTGGGTGGA	2020
Db	1598	TCCCCACCTATCAGGAAAGGTGGAGATGTGAGAGAGTGAAGAGTCTGTGGGTGGA	1657
Qy	2021	AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTCTTAAGAAAGATGGAAATGTG	2080
Db	1658	AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTCTTAAGAAAGATGGAAATGTG	1717
Qy	2081	GGAAACCAACTGGCAAAACAAAGATCTTAAGAGAGACAGGGAAGAGAGTAAGGAA	2140
Db	1718	GGAAACCAACTGGCAAAACAAAGATCTTAAGAGAGACAGGGAAGAGAGTAAGGAA	1777
Qy	2141	GGTCATAGTTGGAGATGGAGATGAGAACTCTTGAAGAAATGGTGCAGACTCCGATGAA	2200
Db	1778	GGTCATAGTTGGAGATGGAGATGAGAACTCTTGAAGAAATGGTGCAGACTCCGATGAA	1837
Qy	2201	GATGATAACAGCTTCTCTCAAAACAACTCTCCCAAGACCCCAAGTCTCTGAATTGGTCG	2260

Db	1838	GATGATAACAGCTTCTCTCAAAACAACAATCTCCACAAGAACCCNAGTCTCTGAAATTGGTCG	1897
Qy	2261	AGTTTTGTAGACAACACCTTTTGTGAGAATTCACCTACTCAGAAATCGAAATCCCAAGAT	2320
Db	1898	AGTTTTGTAGACAACACCTTTTGTGAGAATTCACCTACTCAGAAATCGAAATCCCAAGAT	1957
Qy	2321	GTGGAACCTCTGGAGGGAGAGTGTGCAAGAGCTCTCTGTGGAGAAACAGATAAAGAGA	2380
Db	1958	GTGGAACCTCTGGAGGGAGAGTGTGCAAGAGCTCTCTGTGGAGAAACAGATAAAGAGA	2017
Qy	2381	AATCGGTATTATGATGAGGATGAGGATGAGAGTGAACAATTGCAATTGATCTGGGCCCTT	2440
Db	2018	AATCGGTATTATGATGAGGATGAGGATGAGAGTGAACAATTGCAATTGATCTGGGCCCTT	2077
Qy	2441	AAATTCATGTTAGTGTAGCCAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATAAGCAG	2500
Db	2078	AAATTCATGTTAGTGTAGCCAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATAAGCAG	2137
Qy	2501	GTATCCCGAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTCTTAA	2560
Db	2138	GTATCCCGAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTCTTAA	2197
Qy	2561	AATCAAAAAC 2570	
Db	2198	AATCAAAAAC 2207	

Search completed: January 6, 2004, 13:51:03
Job time : 13119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:49:06 ; Search time 903 Seconds
(without alignments)
11075.767 Million cell updates/sec

Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcgcaggacgttagtg.....atttgaaaaa.....
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2532756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3705	100.0	3705 21	Sequence encoding
2	3688.4	99.6	3711 21	Human ORFX ORF2698
3	3650.4	98.8	3664 23	Human transcriptio
4	3568.8	96.3	3650 22	Human EPLIN (epith
5	3324.4	89.7	3543 22	Human EPLIN (epith
6	2797.6	75.5	2905 22	Human polynucleoti
7	2789.2	75.3	3465 23	DNA encoding novel
8	2736.4	73.9	2749 24	Human cDNA encodin

9	2696.6	72.8	2783	20	AA233566	Human breast tumou
10	2208.8	59.6	2267	21	AAC59489	Human secreted pro
11	2083.6	56.2	2158	21	AAC93483	Human secreted pro
12	1721.4	46.5	1754	24	AB198973	Human cancer suppr
13	1595	43.0	1713	22	AAI59955	Human polynucleoti
14	1568.4	42.3	2207	22	AAH18125	Human cDNA sequenc
15	1433	38.7	1567	21	AAI85888	cDNA encoding huma
16	722	19.5	732	21	AAC98944	Human pancreatic c
17	637.4	17.2	641	20	AA252892	Human prostate tum
18	590.6	15.9	698	22	AAH72864	Human cervical can
19	551	14.9	565	24	ABV98930	Human pancreatic c
20	467.4	12.6	1014	23	AA888277	DNA encoding novel
21	425	11.5	463	24	ABQ58707	Human colon cancer
22	416	11.2	532	22	AAH12979	Human cDNA clone (
23	377.6	10.2	389	21	AAC02079	Human secreted pro
24	362	9.8	586	21	AAA16134	Human colon cancer
25	352.6	9.5	413	25	ACA56082	Chinese hamster si
26	343	9.3	343	24	ABK29495	Colon adenocarcino
27	343	9.3	343	24	ABK29810	Human colon cancer
28	322.2	8.7	382	24	ABQ58853	Human secreted pro
29	321	8.7	376	23	AA572739	DNA encoding novel
30	294.6	8.0	577	23	ABV49483	Human prostate exp
31	279.8	7.6	283	24	ABV88456	Human prostate can
32	277.4	7.5	279	25	ACA55895	Human colon cancer
33	230	6.2	419	22	ABA08387	Chinese hamster si
34	224	6.0	238	25	ACA55583	Human Ca2+ pump PM
35	217.2	5.9	408	23	AA572740	Chinese hamster si
36	212	5.7	247	22	AAH69368	DNA encoding novel
37	211.6	5.7	2226	23	AA572745	Human cervical can
38	210.8	5.7	231	22	AAH72234	DNA encoding novel
39	210.8	5.7	235	22	AAH72170	Human cervical can
40	209.4	5.7	265	22	AAF98667	Human ovarian can
41	199.2	5.4	296	21	AAC03485	Human secreted pro
42	197.8	5.3	232	22	AAI19758	Human breast cance
43	188.2	5.1	269	22	AAI10178	Human breast tumou
44	175.2	4.7	178	24	ABU38497	Human colon tumour
45	141.2	3.8	2336	22	AA533341	DNA encoding human

ALIGNMENTS

RESULT 1
AAAS3826
ID AAAS3826 standard; DNA; 3705 BP.
XX
AC AAAS3826;
XX
DT 03-JAN-2001 (first entry)
XX
DE Sequence encoding lipid associated protein (LIPAP) 2766980CBI.
XX
KW Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
KW antagonist; antibody; cardiovascular disease; neurological disease;
KW gastrointestinal disease; lipid metabolism; detection;
KW amplification; monitoring; hybridisation; antisense; triplex;
KW ribozyme; screening; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 137..2416
FT /*tag= a
FT /product= Lipid associated protein
XX
XX WO200049043-A2.
XX
XX 24-AUG-2000.
PD
XX
XX 18-FEB-2000; 2000WO-US0416Q.
XX
XX 19-FEB-1999; 99US-0120703.
PR 08-JUL-1999; 99US-0142762.

INCYTE PHARM INC.

(INCY-) INCYTE PHARM INC;

Tang YT, Hillman JL, Yue H, Azimzai I, Boegmans M,
WPI; 2000-549264/50.
P-DSNB: AA97286.

P-PSDB; AA97286.

New human lipid-associated proteins, nucleic acids, and antibodies, useful for diagnosis, treatment and prevention of e.g. cardiovascular disease

07 09. 9300: English.

Claim 4: Page 87-88: 93pp, 93a-93g

Lipid-associated proteins (LIPAP) can be used for treating or preventing disorders associated with decreased expression of LIPAP, for screening for agonists or antagonists of LIPAP, and to raise specific antibodies. Antagonists and antagonists of LIPAP are useful for treating diseases associated with reduced or increased levels of LIPAP, e.g. cardiovascular, neurological and gastrointestinal diseases and disorders of lipid metabolism. Fragments of the nucleic acid encoding LIPAP are useful for detection of full length coding sequences, in hybridization and/or amplification assays or for diagnosis or monitoring. Nucleotides encoding LIPAP are used to screen for compounds that specifically modify LIPAP expression, for recombinant production of LIPAP, in gene therapy, as a source of therapeutic antisense, triplex-forming, or ribozyme agents and for genomic mapping. Antibodies to the proteins are used for diagnosis and monitoring of LIPAP-associated disease by immunoassay, and antagonists, in competitive drug screens and for affinity maturation of natural LIPAP.

purification of nucleic acids

Sequence	3705 BP:	1217 A:	743 C:	826 G:	919 T:	938 GAGG
Query Match	100.0 %;	Score 3705;	DB 21:	Length 3705;		
Best Local Similarity	100.0 %;	Pred. No. 0;				
Matches 3705;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
1	GGCGCAGGACGACTAGGTGTTAGCACCTTGGTCGCGACAGGTGGCGCTAGGTAGAGCGCC	60				
1	GGCGCCAGGACGACTAGGTGTTAGCACCTTGGTCGCGACAGGTGGCGCTAGGTAGAGCGCC	60				
61	GGGACCTGTGACGGCTGGTTAGCAGCGCAGAGGAAAGCGCGCTTTTACCGAGGTAATTC	120				
61	GGGACCTGTGACGGCTGGTTAGCAGCGCAGAGGAAAGCGCGCTTTTACCGAGGTAATTC	120				
121	AGTCTCTGTAGACAAGATGGAAATCATCTCCATTTAATAGCGGCAATGGACCTCACTATC	180				
121	AGTGTCTGTAGACACAGATGGAAATCATCTCCATTTAATAGCGGCAATGGACCTCACTATC	180				
181	ATTGAGGGTTAAACAGCCAAAGAACTTCTCTTGTCAACAAGAACAAAGTCATCGGCTATTGT	240				
181	ATTGAGGGTTAAACAGCCAAAGAACTTCTCTTGTCAACAAGAACAAAGTCATCGGCTATTGT	240				
241	GGAAATATTTCTCAAGTACCAGAAAGCGCTGAAGAAACAAACATGGAGAGAGAGAGAG	300				
241	GGAAATATTTCTCAAGTACCAGAAAGCGCTGAAGAAACAAACATGGAGAGAGAGAGAG	300				
301	TAAACACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTTGACTGTGTTAAAGAGAA	360				
301	TAAACACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTTGACTGTGTTAAAGAGAA	360				
361	GTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACAGACTCTCTACGGAAACAGCAGCAC	420				
361	GTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACAGACTCTCTACGGAAACAGCAGCAC	420				
421	TGAGATTAGGCACAGCAGACCACTCTCTCTGCTGAAGTGACAAAGCCACGCTCTTCGG	480				
421	TGAGATTAGGCACAGCAGACCACTCTCTCTGCTGAAGTGACAAAGCCACGCTCTTCGG	480				
481	AGCCAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGA	540				
481	AGCCAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGA	540				

QY 1621 TGCCCTTATTGCTAAGGTGGGTGTCCTGGCTGCAAGTATGGAAGCCAAAGGCCCTCTCTCA 1680
DB |||||
1621 TGCCCTTATTGCTAAGGTGGGTGTCCTGGCTGCAAGTATGGAAGCCAAAGGCCCTCTCTCA 1680
QY 1681 GCAGGAGAAGGAACAGAACCCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCCC 1740
DB |||||
1681 GCAGGAGAAGGAACAGAACCCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCCC 1740
QY 1741 CACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAGAGCCCAA 1800
DB |||||
1741 CACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAGAGCCCAA 1800
QY 1801 ATGGCCCTCTCAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCT 1860
DB |||||
1801 ATGGCCCTCTCAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCT 1860
QY 1861 GAAGAACTAAGACGATCTTCTTCACTGAAGAGAAAGAGCCGCCCATTCACCTGACGAGC 1920
DB |||||
1861 GAAGAACTAAGACGATCTTCTTCACTGAAGAGAAAGAGCCGCCCATTCACCTGACGAGC 1920
QY 1921 TTCAATTTCAAGCACCTCTCTCAAGAGCCCAAAATCTGTCTCCCACTATCAGGAAGG 1980
DB |||||
1921 TTCAATTTCAAGCACCTCTCTCAAGAGCCCAAAATCTGTCTCCCACTATCAGGAAGG 1980
QY 1981 CTGGAGCATCTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCCAGAAAGGAACA 2040
DB |||||
1981 CTGGAGCATCTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCCAGAAAGGAACA 2040
QY 2041 AGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAA 2100
DB |||||
2041 AGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAA 2100
QY 2101 CAAAGAACTTAAAGGAGAGACAGGGAAGAGAAATGAAGAAAGTCTAGTTTGGAGATGGA 2160
DB |||||
2101 CAAAGAACTTAAAGGAGAGACAGGGAAGAGAAATGAAGAAAGTCTAGTTTGGAGATGGA 2160
QY 2161 GAATGAGAACTTGTAGAAATGCTGAGAACTCGATGCGATGAGATGATAACAGCTTCTCTAA 2220
DB |||||
2161 GAATGAGAACTTGTAGAAATGCTGAGAACTCGATGCGATGAGATGATAACAGCTTCTCTAA 2220
QY 2221 ACAACAATCTCCCAAGAACCCCAAGTCTCTGAATTTGGTTCAGATTTGTAGACAAACCTT 2280
DB |||||
2221 ACAACAATCTCCCAAGAACCCCAAGTCTCTGAATTTGGTTCAGATTTGTAGACAAACCTT 2280
QY 2281 TGCTGAAGAAATTCACCTACTCAGAAATCCAGGATGCGAACTCTGGAGGGGAGA 2340
DB |||||
2281 TGCTGAAGAAATTCACCTACTCAGAAATCCAGGATGCGAACTCTGGAGGGGAGA 2340
QY 2341 AGTGGTCAAGAGCTCTCTGTGAAGAAACAGATAAAGAGAAATCGGTATTATGATGAGGA 2400
DB |||||
2341 AGTGGTCAAGAGCTCTCTGTGAAGAAACAGATAAAGAGAAATCGGTATTATGATGAGGA 2400
QY 2401 TGAGGATGAAGAGTGAACAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTTAGC 2460
DB |||||
2401 TGAGGATGAAGAGTGAACAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTTAGC 2460
QY 2461 GAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAGGTATCCCAAGCATGAAATGTA 2520
DB |||||
2461 GAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAGGTATCCCAAGCATGAAATGTA 2520
QY 2521 ATTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAATTCAAAAACAACAAACAA 2580
DB |||||
2521 ATTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAATTCAAAAACAACAAACAA 2580
QY 2581 AACACAAAAACACATCTTAAATCTAGAGATAACCTTAAATTCCTTAAATTCAAAAACAACAA 2640
DB |||||
2581 AACACAAAAACACATCTTAAATCTAGAGATAACCTTAAATTCCTTAAATTCAAAAACAACAA 2640
QY 2641 GTGATGATATGCAATAGTGTGTAAGGCTTGAACCTGGGGAATATTCACACCTGATAATA 2700
DB |||||
2641 GTGATGATATGCAATAGTGTGTAAGGCTTGAACCTGGGGAATATTCACACCTGATAATA 2700
QY 2701 GCCAGATTCTACTGTATTCCCAAAAGGCAATTAAGGTAGATAGATTAGTAGTAGTAT 2760

RESULT 2
AAC77143
ID AAC77143 standard; cDNA; 3711 bp.
XX
AC AAC77143;

DB 2701 GCCAGATTCTACTGTATTTCCAAAAGGCAATATTAAAGTAGATAGATTAGTAGTAT 2760
QY 2761 ATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTTAAACAT 2820
DB |||||
2761 ATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTTAAACAT 2820
QY 2821 TACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAA 2880
DB |||||
2821 TACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAA 2880
QY 2881 TTTAAATTTTGTAGTATTTACCTGTTAAGAGATATTATTAGTCTTTAAATTTTGTAGTTAA 2940
DB |||||
2881 TTTAAATTTTGTAGTATTTACCTGTTAAGAGATATTATTAGTCTTTAAATTTTGTAGTTAA 2940
QY 2941 TTTCTCTGCTGTATATATGAGGAATTTACTACTTTATGTCCTGCTCTCTAAACTACA 3000
DB |||||
2941 TTTCTCTGCTGTATATATGAGGAATTTACTACTTTATGTCCTGCTCTCTAAACTACA 3000
QY 3001 TCCTGAATCGACGCTCTGAGGTATATAACACAGAGCACCTTTTGGAGCAATTTGAAAAA 3060
DB |||||
3001 TCCTGAATCGACGCTCTGAGGTATATAACACAGAGCACCTTTTGGAGCAATTTGAAAAA 3060
QY 3061 CCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGTATC 3120
DB |||||
3061 CCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGTATC 3120
QY 3121 TGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTCTGGTGATATCTGTGCTTC 3180
DB |||||
3121 TGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTCTGGTGATATCTGTGCTTC 3180
QY 3181 TCATAAATTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCCCCCATCT 3240
DB |||||
3181 TCATAAATTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCCCCCATCT 3240
QY 3241 TCCGTGTTAGAGCAAAAGTGAAGAGTTTAAAGAGAGAAAGAAAGAACTGTCTTACACCA 3300
DB |||||
3241 TCCGTGTTAGAGCAAAAGTGAAGAGTTTAAAGAGAGAAAGAAAGAACTGTCTTACACCA 3300
QY 3301 CTGAGCTCAGACCTCTAAACCCCTGATTTCCCTCTATGATGTCCTTTTGTAGACACTA 3360
DB |||||
3301 CTGAGCTCAGACCTCTAAACCCCTGATA TTCCCTCTATGATGTCCTTTTGTAGACACTA 3360
QY 3361 ATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCACAGTATTTCTCAGGGTG 3420
DB |||||
3361 ATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCACAGTATTTCTCAGGGTG 3420
QY 3421 AAATTAACCAACTATAGGCCCTTTTCTTGGAGTATTTTCTAGTCTTAAAGTTTGGGA 3480
DB |||||
3421 AAATTAACCAACTATAGGCCCTTTTCTTGGAGTATTTTCTAGTCTTAAAGTTTGGGA 3480
QY 3481 CATTAATAACTTGAAGTACATTTGTTACACAGTGTATTTCCAAATTTGTATGATGGA 3540
DB |||||
3481 CATTAATAACTTGAAGTACATTTGTTACACAGTGTATTTCCAAATTTGTATGATGGA 3540
QY 3541 GGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTCTAGTCTTACGCAATTTATAGAGATTTAGCTT 3600
DB |||||
3541 GGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTCTAGTCTTACGCAATTTATAGAGATTTAGCTT 3600
QY 3601 AATATTTTGAAGATGTAAACCATCTGCTTTCTTCTAGTCTTACGCAATTTAGCTGAAACATTT 3660
DB |||||
3601 AATATTTTGAAGATGTAAACCATCTGCTTTCTTCTAGTCTTACGCAATTTAGCTGAAACATTT 3660
QY 3661 TTATTCATAAAGATTTTAAATTAATAATTTGAAAAAATTTGAAAAAATTTGAAAAAATTT 3705
DB |||||
3661 TTATTCATAAAGATTTTAAATTAATAATTTGAAAAAATTTGAAAAAATTTGAAAAAATTT 3705

08-FEB-2001 (first entry)

Human ORFX ORF2698 polynucleotide sequence SEQ ID NO:5395.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antiparasitic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antihypertensive; antitumor; cancer; proliferative disorder; hypertension; antianemic; gene therapy; osteoarthritis; hypothyroidism; SCID; AIDS; neurodegenerative disorder; diabetes mellitus; erythematous; infection; cardiovascular disease; systemic lupus erythematosus; burn; wound; cholesterol ester storage; severe combined immunodeficiency; malaria; autoimmune disease; severe combined immunodeficiency; nocturnal haemoglobinuria; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

MO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI: 2000-602362/57.

P-PSDB; AAB42934.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 5; Page 4578-4580; 5507pp; English.

AAC74446 to AAC7605 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antiparasitic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antirheumatic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antihypertensive; antitumor; cancer; proliferative disorder; hypertension; antianemic; gene therapy; osteoarthritis; hypothyroidism; SCID; AIDS; AIDS; viral; erythematous; severe combined immunodeficiency (SCID); AIDS; viral; bacterial or fungal infection; malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 3711 BP; 1215 A; 747 C; 830 G; 919 T; 0 other;

Query Match 99.6%; Score 3688.4; DB 21; Length 3711;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3702; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY	1	GGCCGACGAGCAGTAGGTGTTAGCAGCTTGGTGGGACAGGTGGCTTAGGTAGAGCCG	60
DB	6	GGCCGACGAGCAGTAGGTGTTAGCAGCTTGGTGGGACAGGTGGCTTAGGTAGAGCCG	65
QY	61	GGGACCTGTGACAGGCTGGTAGCAGCGCAGAGGAAAGCGGCTTTTAGCCAGGTATTTC	120
DB	66	GGGACCTGTGACAGGCTGGTAGCAGCGCAGAGGAAAGCGGCTTTTAGCCAGGTATTTC	125
QY	121	AGTGTCTGTAGACAGATGAATCATCTCTTCAATTAATAGACGGCAATGGACCTCCTATC	180
DB	126	AGTGTCTGTAGACAGATGAATCATCTCTTCAATTAATAGACGGCAATGGACCTCCTATC	185
QY	181	ATTGAGGCTTAACGACCAAGAACTTTCTCTTCAACAGAACTGCTCATCGGCTATTGT	240
DB	186	ATTGAGGCTTAACGACCAAGAACTTTCTCTTCAACAGAACTGCTCATCGGCTATTGT	245
QY	241	GGAAATATTCTCCAAAGTACAGAAAGCAGTGAAGAAACAAACATGGAGAAAGAGAAAG	300
DB	246	GGAAATATTCTCCAAAGTACAGAAAGCAGTGAAGAAACAAACATGGAGAAAGAGAAAG	305
QY	301	TAAACACGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTTAAAGAGAA	360
DB	306	TAAACACGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTTAAAGAGAA	365
QY	361	GTGGGAGACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAACAGCAGCAC	420
DB	366	GTGGGAGACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAACAGCAGCAC	425
QY	421	TGAGATTAGGCACAGAGCAGACCATCTCTCTGCTGAAGTGACAAGCCACGCTCTCTGG	480
DB	426	TGAGATTAGGCACAGAGCAGACCATCTCTCTGCTGAAGTGACAAGCCACGCTCTCTGG	485
QY	481	AGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTTAGACTCAGGTCACTCTCTGA	540
DB	486	AGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTTAGACTCAGGTCACTCTCTGA	545
QY	541	AGCCCTGTTCCAGGTCGATATCCACATCAAGGCGGTGAGGATCTTAAAGACCACTC	600
DB	546	AGCCCTGTTCCAGGTCGATATCCACATCAAGGCGGTGAGGATCTTAAAGACCACTC	605
QY	601	AACAGAAAGTAAAGAAATGGAAATTTGTTAGGAGAAATCCAGGCGATGAAGTAGAAAAATC	660
DB	606	AACAGAAAGTAAAGAAATGGAAATTTGTTAGGAGAAATCCAGGCGATGAAGTAGAAAAATC	665
QY	661	AGAAATCAGTGAACACAGATGCTTCCGGGCAAAATAGAGAAATATAATGTTCCGCTGAA	720
DB	666	AGAAATCAGTGAACACAGATGCTTCCGGGCAAAATAGAGAAATATAATGTTCCGCTGAA	725
QY	721	CAGGCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACTAAGATTCTCCGGGCCA	780
DB	726	CAGGCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACTAAGATTCTCCGGGCCA	785
QY	781	AAGCCGAAGTGAAGTGGAAGGAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGA	840
DB	786	AAGCCGAAGTGAAGTGGAAGGAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGA	845
QY	841	AATAGGCCAGGTCAGTTGCTCTCTTACATTTGACTCGGAGAAATAAGAGTAGAGACG	900
DB	846	AATAGGCCAGGTCAGTTGCTCTCTTACATTTGACTCGGAGAAATAAGAGTAGAGACG	905
QY	901	AAATCTGGAATCTCCAGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCA	960
DB	906	AAATCTGGAATCTCCAGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCA	965
QY	961	GGCAGCTGTGTCGAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGG	1020
DB	966	GGCAGCTGTGTCGAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGG	1025
QY	1021	TGCGGAAATCAAAATTCATAAATGGAGCAAAAGAGATGTCGCCCCAGGTCCTGAGGT	1080
DB	1026	TGCGGAAATCAAAATTCATAAATGGAGCAAAAGAGATGTCGCCCCAGGTCCTGAGGT	1085

QY 1081 CTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCG 1140
Db 1086 CTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCG 1145
QY 1141 TTCCACCCCTGCCGAAGATCACTCCC...GTGACTCCCAGGTTAAGAGTGAGGTTCAACA 1197
Db 1146 TTCCACCCCTGCCGAAGATCACTCCCAGGTTAAGAGTGAGGTTCAACA 1205
QY 1198 GCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTC 1257
Db 1206 GCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTC 1265
QY 1258 TCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTC 1317
Db 1266 TCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTC 1325
QY 1318 GAAGACAGTCTATCCAAATGAGAGGCTCTCTTGGCAACCCAGCAGGTGTTCACATCAGCTG 1377
Db 1326 GAAGACAGTCTATCCAAATGAGAGGCTCTCTTGGCAACCCAGCAGGTGTTCACATCAGCTG 1385
QY 1378 CTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTACATGG 1437
Db 1386 CTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTACATGG 1445
QY 1438 AAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGCAACTATCATGA 1497
Db 1446 AAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGCAACTATCATGA 1505
QY 1498 AGGCTTTGGGCACAGACACACAAGGATCTATGGGCAAGCAAAATGAAAAAGAGAT 1557
Db 1506 AGGCTTTGGGCACAGACACACAAGGATCTATGGGCAAGCAAAATGAAAAAGAGAT 1565
QY 1558 TTTGGAGAGACCGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCCAAGGGTAG 1617
Db 1566 TTTGGAGAGACCGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCCAAGGGTAG 1625
QY 1618 AGATGCCCTATTGCTAAGGTGGTGTCTGGTGCAGATATGGAAGCCAAAGCCTCTC 1677
Db 1626 AGATGCCCTATTGCTAAGGTGGTGTCTGGTGCAGATATGGAAGCCAAAGCCTCTC 1685
QY 1678 TCAGCAGGAGAGAGACAGACAGCCAGCTGAACCAAGAGCTGAGATCGCCTGGCCAC 1737
Db 1686 TCAGCAGGAGAGAGAGACAGCCAGCTGNAACCAAGAGCTGAGGATCGCCTGGCCAC 1745
QY 1738 CCCACTGAACTTTGGAAGTTTCAGGAAGTGCTTGGAGGAAGGATCAAAATGTCAAAGCC 1797
Db 1746 CCCACTGAACTTTGGAAGTTTCAGGAAGTGCTTGGAGGAAGGATCAAAATGTCAAAGCC 1805
QY 1798 CAAATGGCCTCTCGAAGACCGAAATCAGCAAGCCGAAAGTTCTTGAGGATGTGATCTAGA 1857
Db 1806 CAAATGGCCTCTCGAAGACCGAAATCAGCAAGCCGAAAGTTCTTGAGGATGTGATCTAGA 1865
QY 1858 TCTGAAGAGCTAAGAGCATCTTCTCACTGAAGGAAGAGCCGCCAATCACTGTAGC 1917
Db 1866-TCTGAAGAGCTAAGAGCATCTTCTCACTGAAGGAAGAGCCGCCAATCACTGTAGC 1925
QY 1918 AGCTTCATTTCAAAGCACCTCTGTCAGAGCCCAAAAATGTGTCCCCACCTATCAGGA 1977
Db 1926 AGCTTCATTTCAAAGCACCTCTGTCAGAGCCCAAAAATGTGTCCCCACCTATCAGGA 1985
QY 1978 AGCCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTGCAGAAAGGAA 2037
Db 1986 AGCCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTGCAGAAAGGAA 2045
QY 2038 ACNAGTGGAAATGCCAAGCTTCTAAGAGAAATGGGAATGTGGGAAAAAACCTGGCA 2097
Db 2046 ACNAGTGGAAATGCCAAGCTTCTAAGAGAAATGGGAATGTGGGAAAAAACCTGGCA 2105
QY 2098 AACAAGAGATCTAAGAGAGACAGGAGAGAGTGAAGAGTCAATAGTTGGAGAT 2157
Db 2106 AACAAGAGATCTAAGAGAGACAGGAGAGAGTGAAGAGTCAATAGTTGGAGAT 2165
QY 2158 GGAGAATGAGAATCTTTGTAGNAATTTGGTGCAGACTCCGATGAAGATGATPACAGCTTCT 2217

Db 2166 GGAGAATGAGAATCTTTGTAGAAAATTTGGTGCAGACTCCGATGAAGATGATAACAGCTTCT 2225
QY 2218 CAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTTGTAGACAACAC 2277
Db 2226 CAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTTGTAGACAACAC 2285
QY 2278 CTTTCTGTAAGAAATTCACACTACTCAGAAATCAGAAATCCCAAGATGTGGAACTCTGGAGGG 2337
Db 2286 CTTTCTGTAAGAAATTCACACTACTCAGAAATCAGAAATCCCAAGATGTGGAACTCTGGAGGG 2345
QY 2338 AGAAGTGTTCAAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGTATTATGATGA 2397
Db 2346 AGAAGTGTTCAAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGTATTATGATGA 2405
QY 2398 GGATCAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCTTAAATTCATGTTAGTGT 2457
Db 2406 GGATCAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCTTAAATTCATGTTAGTGT 2465
QY 2458 AGCAGGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTATCCAGCATGAAAT 2517
Db 2466 AGCAGGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTATCCAGCATGAAAT 2525
QY 2518 GTAAATTTACTTTGGAAAGTAACCTTTGGAAAAGAAATTCCTTCTTAAATCAAAAACAA 2577
Db 2526 GTAAATTTACTTTGGAAAGTAACCTTTGGAAAAGAAATTCCTTCTTAAATCAAAAACAA 2585
QY 2578 AAAAAACAAAAACACACATCTTAATCTAGAGATACTTTACTTAAATTTCTTCAATTTA 2637
Db 2586 AAAAAACAAAAACACACATCTTAATCTAGAGATACTTTACTTAAATTTCTTCAATTTA 2645
QY 2638 GCAGTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2697
Db 2646 GCAGTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2705
QY 2698 ATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATGATGATGAT 2757
Db 2706 ATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATGATGATGAT 2765
QY 2758 TATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGGCTTAAA 2817
Db 2766 TATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGGCTTAAA 2825
QY 2818 CATTACGACTGAATGCACTTTTAGTATTAAGGGCAGAGTTTCTATATTTTAAATGAATAC 2877
Db 2826 CATTACGACTGAATGCACTTTTAGTATTAAGGGCAGAGTTTCTATATTTTAAATGAATAC 2885
QY 2878 CAATTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTTACTCTTTTAAATTTTGTAGT 2937
Db 2886 CAATTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTTACTCTTTTAAATTTTGTAGT 2945
QY 2938 TAAATTTCTGCTGTGATATATGAGAAATTTACTATTTTATGCTGCTCTCTAAACT 2997
Db 2946 TAAATTTCTGCTGTGATATATGAGAAATTTACTATTTTATGCTGCTCTCTAAACT 3005
QY 2998 ACATCTCGAACTCCAGCTCTGAGGTATATACAGAGAGACACTTTTGGAGCAATGAA 3057
Db 3006 ACATCTCGAACTCCAGCTCTGAGGTATATACAGAGAGACACTTTTGGAGCAATGAA 3065
QY 3058 AAAACCAACTACACTCTTTGGGTGTTAGAGAGATCTGCTGCTCCCAAAATGAAGCTTTTGT 3117
Db 3066 AAAACCAACTACACTCTTTGGGTGTTAGAGAGATCTGCTGCTCCCAAAATGAAGCTTTTGT 3125
QY 3118 ATCTGCCAGTGAATTTACTGTACTCCAAATGAATTTCTTTTCTGGTGATATCTGTGC 3177
Db 3126 ATCTGCCAGTGAATTTACTGTACTCCAAATGAATTTCTTTTCTGGTGATATCTGTGC 3185
QY 3178 TTCTCATTAATTTACTGAAAGTGCATAATTTTGTAAATACCTTCGGGATCACTGTCCCCCA 3237
Db 3186 TTCTCATTAATTTACTGAAAGTGCATAATTTTGTAAATACCTTCGGGATCACTGTCCCCCA 3245
QY 3238 TCTTCCGTGTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTTACA 3297

QY 1004 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAAATGTG 1063
DB 961 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAAATGTG 1020
QY 1064 CCCCCAGGTCTCTGAGTCTGCATCACCCATCAGCAAGGGGAAAGATTTCTGCAAAATGAG 1123
DB 1021 CCCCCAGGTCTCTGAGTCTGCATCACCCATCAGCAAGGGGAAAGATTTCTGCAAAATGAG 1080
QY 1124 AATAGCTGCGAGTCCGTTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCGAGGTTAAG 1183
DB 1081 AATAGCTGCGAGTCCGTTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCGAGGTTAAG 1140
QY 1184 AGTGAGGTTCAACAGCGCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGGCTCCAGT 1243
DB 1141 AGTGAGGTTCAACAGCGCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGGCTCCAGT 1200
QY 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGACACTGCAAGAGAGACC 1303
DB 1201 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGACACTGCAAGAGAGACC 1260
QY 1304 TGGTGGAAATGTGAGAAGACAGTCTATCCAAATGGAGGCTCTTTGGCCAAACAGCAGGTG 1363
DB 1261 TGGTGGAAATGTGAGAAGACAGTCTATCCAAATGGAGGCTCTTTGGCCAAACAGCAGGTG 1320
QY 1364 TTTCACATCAGCTGCTTCCGTTGCTCTCTATTGCAACAACAACTCAGTCTAGGAACATAT 1423
DB 1321 TTTCACATCAGCTGCTTCCGTTGCTCTCTATTGCAACAACAACTCAGTCTAGGAACATAT 1380
QY 1424 GCATCTTTACATGGAAGAAATCTATTGTGAAGCCTCACTTCAATCAACTCTTTAAATCTAAG 1483
DB 1381 GCATCTTTACATGGAAGAAATCTATTGTGAAGCCTCACTTCAATCAACTCTTTAAATCTAAG 1440
QY 1484 GGCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAAAT 1543
DB 1441 GGCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAAAT 1500
QY 1544 GAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGAGCCCTCAC 1603
DB 1501 GAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGAGCCCTCAC 1560
QY 1604 AGCCAGGGGTAGAAGATGCCCTATTGTGAAGTGGGTGCTCTGCTGCTCAAGTATGGA 1663
DB 1561 AGCCAGGGGTAGAAGATGCCCTATTGTGAAGTGGGTGCTCTGCTGCTCAAGTATGGA 1620
QY 1664 GCCAAGGCTCTCTCAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGG 1723
DB 1621 GCCAAGGCTCTCTCAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGG 1680
QY 1724 ATGCGCTGGCCACCCCCCACTGAACTTGAAGTTCAGGAAGTGCCTTGAGGAAGGATC 1783
DB 1681 ATGCGCTGGCCACCCCCCACTGAACTTGAAGTTCAGGAAGTGCCTTGAGGAAGGATC 1740
QY 1784 ABAATGTCAAGGCCAAATGGCTCTCTGAAGAGCAATCAGCAGCCCGAAGTCTCTGAG 1843
DB 1741 ABAATGTCAAGGCCAAATGGCTCTCTGAAGAGCAATCAGCAGCCCGAAGTCTCTGAG 1800
QY 1844 GATGTGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAGAAAGAGCCGC 1903
DB 1801 GATGTGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAGAAAGAGCCGC 1860
QY 1904 CCATTCACTGTAGCAGCTTCATTTCAAAGCACTCTCTCAAGAGGCCCAAAACTGTGTC 1963
DB 1861 CCATTCACTGTAGCAGCTTCATTTCAAAGCACTCTCTCAAGAGGCCCAAAACTGTGTC 1920
QY 1964 CCACCTATCAGNAAGGCTGGAGCATGTGAGAGCAGAGTGAAGATCTGTGGTGGGAAGA 2023
DB 1921 CCACCTATCAGNAAGGCTGGAGCATGTGAGAGCAGAGTGAAGATCTGTGGTGGGAAGA 1980
QY 2024 GTTGCAAGAAAGGAACAAGTGGAAATGCCAGGCTTCTTAAGAGNAATGGGAATGTGGA 2083
DB 1981 GTTGCAAGAAAGGAACAAGTGGAAATGCCAGGCTTCTTAAGAGNAATGGGAATGTGGA 2040

QY 2084 AAAAAAAGCTGGCAAAAAAAGAAATCTTAAGAGAGACAGGGAAGAGAAATGAGGAGGT 2143
DB 2041 AAAAAAAGCTGGCAAAAAAAGAAATCTTAAGAGAGACAGGGAAGAGAAATGAGGAGGT 2100
QY 2144 CATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGACAGCTCCGATGAAGAT 2203
DB 2101 CATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGACAGCTCCGATGAAGAT 2160
QY 2204 GATAACAGCTTCTCTCAACAAACAAATCTCCACAAAGCCCAAGTCTCTGAAATTTGGTTCGAGT 2263
DB 2161 GATAACAGCTTCTCTCAACAAACAAATCTCCACAAAGCCCAAGTCTCTGAAATTTGGTTCGAGT 2220
QY 2264 TTTGTAGACAACACCTTTGCTGAGAAATTCACACTCAGAAATCAGAAATCCAGGATGTG 2323
DB 2221 TTTGTAGACAACACCTTTGCTGAGAAATTCACACTCAGAAATCAGAAATCCAGGATGTG 2280
QY 2324 GAACTCTGGAGGAGAGAGTGGTCAACAGAGCTCTCTGTGGAAAGAACACAGATAAAGAGAAAT 2383
DB 2281 GAACTCTGGAGGAGAGAGTGGTCAACAGAGCTCTCTGTGGAAAGAACACAGATAAAGAGAAAT 2340
QY 2384 CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTTAAA 2443
DB 2341 CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTTAAA 2400
QY 2444 TTTCATGTTAGTGTAGCCAGGACCTGCCCTTTGTCCAAATGTGATGCACATTAAGCAGGTA 2503
DB 2401 TTTCATGTTAGTGTAGCCAGGACCTGCCCTTTGTCCAAATGTGATGCACATTAAGCAGGTA 2460
QY 2504 TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACCTTCGAAAGAAATTCCTCTTTAAAT 2563
DB 2461 TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACCTTCGAAAGAAATTCCTCTTTAAAT 2520
QY 2564 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAACTTTTACTTTA 2623
DB 2521 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAACTTTTACTTTA 2580
QY 2624 AATTCCTCATTTTACAGTGATGATATGCAATGAGTGTCTTAAGGCTTTGTAAGTGGGAAA 2683
DB 2581 AATTCCTCATTTTACAGTGATGATATGCGTAAGTGCTGTAAAGGCTTTGTAAGTGGGAAA 2640
QY 2684 TATTCACCTCATTAATACCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGA 2743
DB 2641 TATTCACCTCATTAATACCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGA 2700
QY 2744 TAGATGATTAGTATGATATTTCTACACATTTTGGNAATTAGAGAACATACAGAGGAA 2803
DB 2701 TAGATGATTAGTATGATATTTCTACACATTTTGGNAATTAGAGAACATACAGAGGAA 2760
QY 2804 TTTAGGGGCTTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATAT 2863
DB 2761 TTTAGGGGCTTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATAT 2820
QY 2864 TTTTAAATGAATACCAATTTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTATAGTCTT 2923
DB 2821 TTTTAAATGAATACCAATTTTAAATTTTGTAGTATTTTACCTGTTAAGAGATTTATAGTCTT 2880
QY 2924 TAAATTTTGTAGTATTTTCTGCTGTGATATATATGAGGAATTTTACTACTTTATGTC 2983
DB 2881 TAAATTTTGTAGTATTTTCTGCTGTGATATATATGAGGAATTTTACTACTTTATGTC 2940
QY 2984 CTGCTCTCTAAACTACATCTCTGAACTCGAGTCTGAGGTATATATCAACAGAGACACTTT 3043
DB 2941 CTGCTCTCTAAACTACATCTCTGAACTCGAGTCTGAGGTATATATCAACAGAGACACTTT 3000
QY 3044 TTGAGGCAATTTGAAAAACCAACCTACACTCTTCCGCTGTAGAGAGATCTGCTGCTCCC 3103
DB 3001 TTGAGGCAATTTGAAAAACCAACCTACACTCTTCCGCTGTAGAGAGATCTGCTGCTCCC 3060
QY 3104 AATAAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTTCT 3163
DB 3061 AATAAGCTTTTGTATCTGCGAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTTCT 3120
QY 3164 GGTGATATCTGCTCTCTCAATATTCTGAAGCTGCAATATTTTGTAGTAATACCTTCGGG 3223

Db 489 CTCAGGTCACCTCTGTAAGCCCTCGTTACGGTTCGATATATCCCCACATCAAGAGCGGTGAG 548
Qy 584 GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAATTTGTCTAGGAGATCCAGG 643
Db 549 GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAATTTGTCTAGGAGATCCAGG 608
Qy 644 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTTGGGCAAAATAGAGAAA 703
Db 609 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTTGGGCAAAATAGAGAAA 668
Qy 704 TATTAATGTCGGTGAACAGGCTTAAAGATGATGTTTGAGAAAGGTGAACCAACTCAAAC 763
Db 669 TATAATGTTCCGTTGAACAGGCTTAAAGATGATGTTTGAGAAAGGTGAACCAACTCAAAC 728
Qy 764 AAGATTCCTCGGCCCAAGCCGGAAGTCAAGTGGAGGAAGATCTCTGAAACAGCTAT 823
Db 729 AAGATTCCTCGGCCCAAGCCGGAAGTCAAGTGGAGGAAGATCTCTGAAACAGCTAT 788
Qy 824 TCTCTAGATGACCTTGAATATAGGCCAGGTCAGTTGTCTATCTTCTACATTTGACTCGGAG 883
Db 789 TCTCTAGATGACCTTGAATATAGGCCAGGTCAGTTGTCTATCTTCTACATTTGACTCGGAG 848
Qy 884 AAAAAAGAGAGTAGCAAAATCTGGAACCTTCCAGGCTCTCAGAAACCTCTATAAAGGAT 943
Db 849 AAAAAAGAGAGTAGCAAAATCTGGAACCTTCCAGGCTCTCAGAAACCTCTATAAAGGAT 908
Qy 944 CGAATGSCCAAGTACAGGCGAGTGTGTCCAAACAAAGAGCTCAACCAACTATACAAAT 1003
Db 909 CGAATGSCCAAGTACAGGCGAGTGTGTCCAAACAAAGAGCTCAACCAACTATACAAAT 968
Qy 1004 GAGCTGAAGCCAGTGTGGCGGAAATCAAAATTCATAAAATGAGCAAAAGGAGAGATGTG 1063
Db 969 GAGCTGAAGCCAGTGGTGGCGGAAATCAAAATTCATAAAATGAGCAAAAGGAGAGATGTG 1028
Qy 1064 CCCCCAGGTCCTGAGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1123
Db 1029 CCCCCAGGTCCTGAGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1088
Qy 1124 AATAGCTGSCAGTCCGTTCCAGCCCTGCGGAGATGACTCCC---GTGACTCCCAGGTT 1180
Db 1089 AATAGCTGSCAGTCCGTTCCAGCCCTGCGGAGATGACTCCCAGGTCAGTCCCAGGTT 1148
Qy 1181 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTTCCAGAGCCTCC 1240
Db 1149 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTTCCAGAGCCTCC 1208
Qy 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATCAAGAAAGTTTCAGGCACCTGCAAGAGAG 1300
Db 1209 AGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATCAAGAAAGTTTCAGGCACCTGCAAGAGAG 1268
Qy 1301 ACCTGGTGGAAATGTGAGAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACAGCAG 1360
Db 1269 ACCTGGTGGAAATGTGAGAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACAGCAG 1328
Qy 1361 GTGTTTACATCAGCTGCTTCCGTTGCTCTGCTATTTGCAACCAAACTAGTCTAGGAACA 1420
Db 1329 GTGTTTACATCAGCTGCTTCCGTTGCTCTGCTATTTGCAACCAAACTAGTCTAGGAACA 1388
Qy 1421 TATGCATCTTTTACATGGAAGAAATCTATTTGAGCCCTCAGTTCAATCAACTCTTTAAATCT 1480
Db 1389 TATGCATCTTTTACATGGAAGAAATCTATTTGAGCCCTCAGTTCAATCAACTCTTTAAATCT 1448
Qy 1481 AAGGGCAACTATGATAAGGCTTTGGGCACAGACCAACCAAGGATCTATGGGCAAGCAAA 1540
Db 1449 AAGGGCAACTATGATAAGGCTTTGGGCACAGACCAACCAAGGATCTATGGGCAAGCAAA 1508
Qy 1541 AATGAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCATAATGCAAGGAGACCCCT 1600
Db 1509 AATGAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCATAATGCAAGGAGACCCCT 1568
Qy 1601 CACAGCCCAAGGGTAGAAGATCCCTATTTGCTAGGTGGTGTCTGCTGCTGCAAGATG 1660
Db 1569 CACAGCCCAAGGGTAGAAGATCCCTATTTGCTAAGGTGGTGTCTGCTGCTGCAAGATG 1628

Qy 1661 GAAGCCAAAGGCTCTCTCAGCAGGAGAGGAACACAGCCAGCTGAACCCAGAGCTG 1720
Db 1529 GAAGCCAAAGGCTCTCTCAGCAGGAGAGGAACACAGCCAGCTGAACCCAGAGCTG 1688
Qy 1721 AGGATCGCTCGCCACCCCTCAGCTGAACTTGGAACTTCAGGAAGTGCCTTTGAGGAAGG 1780
Db 1689 AGGATCGCTCGCCACCCCTCAGCTGAACTTGGAACTTCAGGAAGTGCCTTTGAGGAAGG 1748
Qy 1781 ATCAAAATGTCAAAAGCCCAAAATGGCTCTGGAAGACGAAATCAGCAAGCCCAAGATTCCT 1840
Db 1749 ATCAAAATGTCAAAAGCCCAAAATGGCTCTGGAAGACGAAATCAGCAAGCCCAAGATTCCT 1808
Qy 1841 GAGGATGTCGATCTAGATCTGGAAGAGCTAGACGATCTTCTTCACTGAAGAAAGAAAGC 1900
Db 1809 GAGGATGTCGATCTAGATCTGGAAGAGCTAGACGATCTTCTTCACTGAAGAAAGAAAGC 1868
Qy 1901 CCCCCATTCACGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960
Db 1869 CCCCCATTCACGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1928
Qy 1961 TCCCCACCTATCAGGAAAAGGCTGGAGCATGTGAGAGCAGAGTGAAGAGTCTGTGGGTGGA 2020
Db 1929 TCCCCACCTATCAGGAAAAGGCTGGAGCATGTGAGAGCAGAAATGAAGATCTGTGGGTGGA 1988
Qy 2021 AGAGTTGCAGAAAGAAAACAAGTGGAAAATGCCAAGGCTTTCTAAGAGAAATGGGAATGTG 2080
Db 1989 AGAGTTGCAGAAAGAAAACAAGTGGAAAATGCCAAGGCTTTCTAAGAGAAATGGGAATGTG 2048
Qy 2081 GGAAGAAACAACTGGCAAAACAAAGAAATCTAAGAGAGACAGGGAAGAGAAATGAAGAA 2140
Db 2049 GGAAGAAACAACTGGCAAAACAAAGAAATTTAAGAGAGACAGGGAAGAGAAATGAAGAA 2108
Qy 2141 GGTCTAGTCTTGGAGATGGAGAAATGAGAAATCTGTAGAAAATGGTGCAGATCCGATCAA 2200
Db 2109 GGTCTAGTCTTGGAGATGGAGAAATGAGAAATTTGTAGAAAATGGTGCAGATCCGATCAA 2168
Qy 2201 GATGATAACAGCTTCTCTCAAAACAACAAATCTCCCAAGAACCCCAAGTCTCTGAAATGGTGC 2260
Db 2169 GATGATAACAGCTTCTCTCAAAACAACAAATTTCCCAAGAACCCCAAGTCTTCTGAAATGGTGC 2228
Qy 2261 AGTTTGTAGACAACACTTGTGTGAAGAAATCTACTACTCAGAAATCAGAAATCCAGGAT 2320
Db 2229 AGTTTGTAGACAACACTTGTGTGAAGAAATCTACTACTCAGAAATCAGAAATCCAGGAT 2288
Qy 2321 GTGGAACTCTGGGAGGAGAGTGGTCAAAGAGTCTCTGTGGAGAAACAGAAATGAAGAGA 2380
Db 2289 GTGGAACTTTGGGAGGAGAGAGTGGTCAAAGAGTCTCTGTGGAGAAACAGATGAAGAGA 2348
Qy 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATGCAATGATGATGATGATGAT 2440
Db 2349 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATGCAATGATGATGATGATGAT 2408
Qy 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAAGCAG 2500
Db 2409 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAAGCAG 2468
Qy 2501 GTATCCAGGATGAATGTAAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 2560
Db 2469 GTATCCAGGATGAATGTAAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 2528
Qy 2561 AATCAAAAACAAAACAAAACAAAACAAAACAAATCTAAATAGAGATAACTTTTAC 2620
Db 2529 AATCAAAAACAAAACAAAACAAAACAAAACAAATCTAAATAGAGATAACTTTTAC 2588
Qy 2621 TTAATTTCTTCATTTTAGCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2680
Db 2589 TTAATTTCTTCAT-----CAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2643
Qy 2681 AAATATCCACCTGATATATAGCCAGATTTCTACTGTATTCCTCAAAAAGGCAATATTAAAGT 2740
Db 2644 AAATATCCACCTGATATATAGCCAGATTTCTACTGTATTTCCAAAAGGCAATATTAAAGT 2703

2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2800
|||||
2704 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2763
|||||
2801 GAAATTTAGGGCTTAAACATTACGACTGAATGACCTTTAGTATAAGGCGCAGTTTGT 2860
|||||
2764 GAAATTTAGGGCTTAAACATTACGACTGAATGACCTTTAGTATAAGGCGCAGTTTGT 2822
|||||
2861 TATTTTAAATGAATACCAATTTTAAATTTTATTTTACCTGTTTAAAGAGATTAATTTAGT 2920
|||||
2823 TATTTTAAATGAATACCAATTTTAAATTTTATTTTATTTTACCTGTTTAAAGAGATTAATTTAGT 2882
|||||
2921 CTTTAAATTTTATTTAGGTTAAATTTTCTTTGCTGTGATATATATAGGAAATTTACTACTTTAT 2980
|||||
2883 CTTTAAATTTTATTTAGGTTAAATTTTCTTTGCTGTGATATATATAGGAAATTTACTACTTTAT 2942
|||||
2981 GTCTGCTCTCTAAACTACATCTGTAACCTGAACTCGAGTCTGAGGTATATATACACAGAGCAC 3040
|||||
2943 GTCTGCTCTCTAAACTACATCTGTAACCTGAACTCGAGTCTGAGGTATATATACACAGAGCAC 3000
|||||
3041 TTTTGTAGGCAATTTGAAACCAACCTACACTCTTCCGTCCTTAGAGAGATCTGCTGTCT 3100
|||||
3001 TTTTGTAGGCAATTTGAAACCAACCTACACTCTTCCGTCCTTAGAGAGATCTGCTGTCT 3060
|||||
3101 CCCAAATAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTTGCTTTCTTT 3160
|||||
3061 CCCAAATAAGCTTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAATGATTTGCTTTCTTT 3120
|||||
3161 TCTGTGATATCTGTGCTTCTCATATTTACTGAAAGCTGCAATTTTGTAGTAATACCTTC 3220
|||||
3121 TCTGTGATATCTGTGCTTCTCATATTTACTGAAAGCTGCAATTTTGTAGTAATACCTTC 3180
|||||
3221 GGGATCACTGTCCCCCATCTTCCGTTGTAGAGCAAGTGAAGTGTAAAGGAGGAGAA 3280
|||||
3181 GGGATCACTGTCCCCCATCTTCCGTTGTAGAGCAAGTGAAGTGTAAAGGAGGAGAA 3240
|||||
3281 GAAAGAACTGCTTACACACTTTGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTTATGAT 3340
|||||
3241 GAAAGAACTGCTTACACACTTTGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTTATGAT 3300
|||||
3341 GTCCCTTTTGTAGACACTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 3400
|||||
3301 GTCCCTTTTGTAGACACTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 3360
|||||
3401 ATCAGATATTTCTAGGTTGAAATTTAAACCAACTATAGGCTTTTCTTTGGATGATTTT 3460
|||||
3361 ATCAGATATTTCTAGGTTGAAATTTAAACCAACTATAGGCTTTTCTTTGGATGATTTT 3420
|||||
3461 CTAGCTTTAAGTTTGGGACATTTAAACTTGAGTACATTTTGTACACAGTTGATAT 3520
|||||
3421 CTAGCTTTAAGTTTGGGACATTTAAACTTGAGTACATTTTGTACACAGTTGATAT 3480
|||||
3521 TCCAAATTTGATGGATGGGAGGAGGTTGCTTAAAGCTGTAGGCTTTTCTTTGACTGC 3580
|||||
3481 TCCAAATTTGATGGATGGGAGGAGGTTGCTTAAAGCTGTAGGCTTTTCTTTGACTGC 3540
|||||
3581 ATTTATAGAGATTTAGCTTTAATATTTTATTTAGAGATGTAACATCTGCTTTCTTAGTC 3640
|||||
3541 ATTTATAGAGATTTAGCTTTAATATTTTATTTAGAGATGTAACATCTGCTTTCTTAGTC 3600
|||||
3641 TTACCTAGTCTGAAACATTTTATTTCAATAAAGATTTTAAATTTAAATTTG 3690
|||||
3601 TTACCTAGTCTGAAACATTTTATTTCAATAAAGATTTTAAATTTAAATTTG 3650
|||||

RESULT 5
AAF55696 standard; DNA; 3543 BP.
X
X AAF55696;
X
X 11-JUN-2001 (first entry)
X

DE Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform DNA.
XX
KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer; ss.
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 474..2276
FT /tag= a
FT /transl_except= "(pos: 825..827, aa: Val)"
FT /transl_except= "(pos: 1479..1481, aa: Asp)"
FT /transl_except= "(pos: 1627..1629, aa: Lys)"
FT /transl_except= "(pos: 1860..1861, aa: Ser)"
FT /transl_except= "(pos: 1968..1970, aa: Ser)"
FT /transl_except= "(pos: 2031..2033, aa: Ile)"
FT /transl_except= "(pos: 2088..2090, aa: Ser)"
FT /transl_except= "(pos: 2106..2108, aa: Ser)"
FT /product= "EPLIN (epithelial protein lost in neoplasm)"
XX
WO200118019-A1.
PN
XX 15-MAR-2001.
PD
XX 08-SEP-2000; 2000WO-US24689.
XX
XX 08-SEP-1999; 99US-0153024.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Chang DB, Maul RS;
XX
XX WPI: 2001-244555/25.
XX P-PSDB; AAB67700.
XX
XX New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
XX prognostic and therapeutic applications over the course of cell
XX proliferative disorders associated with EPLIN
XX
XX Claim 4; Page 42-43; 59pp; English.
XX
XX The present sequence encodes a human EPLIN (epithelial protein lost in
XX neoplasm)-alpha isoform. The specification also describes EPLIN-beta.
XX EPLIN is a tumour suppressor protein, whose expression is altered in
XX multiple common human tumour types. EPLIN nucleic acids and proteins are
XX used in screening assays to detect molecules that specifically bind to
XX EPLIN nucleic acids, proteins or derivatives and thus have potential use
XX as agonist or antagonist of EPLIN, in particular molecules that affect
XX cell proliferation. Thus the assays are useful for screening molecules
XX with potential utility as anticancer drugs or lead compounds for drug
XX development. EPLIN nucleic acids, proteins are useful for detecting a
XX cell proliferative disorder in a subject. EPLIN polynucleotides are
XX useful in gene therapy techniques. EPLIN is useful as a marker that
XX can be diagnostically, prognostically and therapeutically used over
XX the course of a cell proliferative disorder associated with EPLIN.
XX
XX Sequence 3543 BP; 1152 A; 711 C; 771 G; 909 T; 0 other;

Query Match 89.7%; Score 3324.4; DB 22; Length 3543;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3371; Conservative 0; Mismatches 11; Indels 10; Gaps 3;
QY 302 AACACCGGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAG 361
|||||
DB 159 AACACCGGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAG 218
|||||
QY 362 TGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACACGACACT 421
|||||
DB 219 TGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACACGACACT 278
|||||
QY 422 GAGATTAGGCACAGACAGACCCATCTCTGCTGAAGTGAACGCCACGCTGCTTCTGGA 481
|||||

Db 279 GAGATTAGGCACAGCAGACCCATCTCTCTGCTGAAAGTGAAGCCACGCTGCTTCTGGA 338
Qy 482 GCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAA 541
Db 339 GCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAA 398
Qy 542 GCCCTGTTGAGGTCGATATCCCAATCAAGACCGGTGAGGATCTTAAAGACCACTCA 601
Db 399 GCCCTGTTGAGGTCGATATCCCAATCAAGACCGGTGAGGATCTTAAAGACCACTCA 458
Qy 602 ACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGGATGAAGTAGAAAATCA 661
Db 459 ACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGGATGAAGTAGAAAATCA 518
Qy 662 GAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAAC 721
Db 519 GAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAAC 578
Qy 722 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCCGGGCCAA 781
Db 579 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCCGGGCCAA 638
Qy 782 AGCGAAGTCAAGTGAAGAAAGATCTCTGAACCAAGCTATTTCTTAGATGACCTGGAA 841
Db 639 AGCGAAGTCAAGTGAAGAAAGATCTCTGAACCAAGCTATTTCTTAGATGACCTGGAA 698
Qy 842 ATAGGCCAGGTTCACGCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 901
Db 699 ATAGGCCAGGTTCACGCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 818
Qy 902 AATCTGGAATTCACGCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 961
Db 759 AATCTGGAATTCACGCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 818
Qy 962 GCAGCTGTGTCACCAACAAAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGT 1021
Db 819 GCAGCTGTGTCACCAACAAAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGT 878
Qy 1022 GCGGAATCAAAATTCATAAATGGACCAAGAGAAATGTGCCCCAGGTCTTGAGGTC 1081
Db 879 GCGGAATCAAAATTCATAAATGGACCAAGAGAAATGTGCCCCAGGTCTTGAGGTC 938
Qy 1082 TGCAATCACCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 1141
Db 939 TGCAATCACCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 998
Qy 1142 TCCACCCCTCCGAAGATGACTCCC---GTGACTCCAGGTTAAGAGTGAGGTTCAACAG 1198
Db 999 TCCACCCCTCCGAAGATGACTCCCAGGTTAAGAGTGAGGTTCAACAG 1058
Qy 1199 CTTGTCTATCCCAAGCCACTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTTCT 1258
Db 1059 CTTGTCTATCCCAAGCCACTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTTCT 1118
Qy 1259 CTTCCCAAGCAATGAAGATTTTCAAGGACCTGCAAGAGAGACCTGCGTGAATGTGAG 1318
Db 1119 CTTCCCAAGCAATGAAGATTTTCAAGGACCTGCAAGAGAGACCTGCGTGAATGTGAG 1178
Qy 1319 AAGACAGTCTATCCAAATGGAGGCTCTCTTGGCCAAACCCAGCAGGTGTTTCAATCAGCTGC 1378
Db 1179 AAGACAGTCTATCCAAATGGAGGCTCTCTTGGCCAAACCCAGCAGGTGTTTCAATCAGCTGC 1238
Qy 1379 TTCCGTTGCTCTATTTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTACATGA 1438
Db 1239 TTCCGTTGCTCTATTTGCAACCAACAACTCAGTCTAGGAACATATGCAATCTTTACATGA 1298
Qy 1439 AGAATCTATTGTAAAGCTCACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAA 1498
Db 1299 AGAATCTATTGTAAAGCTCACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAA 1358
Qy 1499 GCGTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACCAAGAGATT 1558
Db 1359 GCGTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACCAAGAGATT 1418

Qy 1559 TTGGAGAGACCAGCCAGCTTGCBAATGCAAGGGAGACCCCTTCACAGCCCGGGGTAGAA 1618
Db 1419 TTGGAGAGACCAGCCAGCTTGCBAATGCAAGGGAGACCCCTTCACAGCCCGGGGTAGAA 1478
Qy 1619 GATGCCCTTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGAAAGCCCAAGCCCTCTCT 1678
Db 1479 AATGCCCTTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGAAAGCCCAAGCCCTCTCT 1538
Qy 1679 CAGCAGGAGAGGAGACCAAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCC 1738
Db 1539 CAGCAGGAGAGGAGACCAAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCC 1598
Qy 1739 CCCACTGAACCTTGGAAAGTTCAGAAAGTCCCTTGGAGGAGGGATCAAAATGTCAAGGCC 1798
Db 1599 CCCACTGAACCTTGGAAAGTTCAGAAAGTCCCTTGGAGGAGGGATCAAAATGTCAAGGCC 1658
Qy 1799 AAATGGCTCTCTGAAGACGAAATCAGCAAGCCGAGTTCTCTGAGGATGTGATCTAGAT 1858
Db 1659 AAATGGCTCTCTGAAGACGAAATCAGCAAGCCGAGTTCTCTGAGGATGTGATCTAGAT 1718
Qy 1859 CTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCAATTCACCTAGCA 1918
Db 1719 CTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCAATTCACCTAGCA 1778
Qy 1919 GCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1978
Db 1779 GCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1838
Qy 1979 GCGTGGAGCATGTGAGAGCAGATGAAGCTGTGTGGGTGGAAGAGTTGCGAAGAGGAAA 2038
Db 1839 GCGTGGAGCATGTGAGAGCAGATGAAGCTGTGTGGGTGGAAGAGTTGCGAAGAGGAAA 1898
Qy 2039 CAAAGTGGAAATGCCCAAGCTTCTAAGAGAAATGGAATGTGGGAAACCAACCTGGCAA 2098
Db 1899 CAAAGTGGAAATGCCCAAGCTTCTAAGAGAAATGGAATGTGGGAAACCAACCTGGCAA 1958
Qy 2099 AACAAAGAAATCTAAGGAGAGACAGGGAAGAAAGTAAAGGAAGTCTAGTTTGGAGATG 2158
Db 1959 AACAAAGAAATCTAAGGAGAGACAGGGAAGAAAGTAAAGGAAGTCTAGTTTGGAGATG 2018
Qy 2159 GAGAAATGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTC 2218
Db 2019 GAGAAATGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTC 2078
Qy 2219 AAACAAACATCTCCACAGAACCCCAAGTCTCTGAATTTGGTCCAGTTTGTAGACAACACC 2278
Db 2079 AAACAAACATTTCCACAGAACCCCAAGTCTTTGAATTTGGTCCAGTTTGTAGACAACACC 2138
Qy 2279 TTTGCTGAAGAAATTCACCTACTCAGAAATCCAGGAATCCAGGATGTGGAACCTCTGGAGGGA 2338
Db 2139 TTTGCTGAAGAAATTCACCTACTCAGAAATCCAGGAATCCAGGATGTGGAACCTTTGGAGGGA 2198
Qy 2339 GAAATGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAATCGGTATTATGATGAG 2398
Db 2199 GAAATGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAATCGGTATTATGATGAG 2258
Qy 2399 GATGAGATGAAGAGTGAACAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTTA 2458
Db 2259 GATGAGATGAAGAGTGAACAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTTA 2318
Qy 2459 GCGAGCCACTGCCCCCTTTGTCAAAATGTGACATGAAGCAGGTATCCAGCATGAAATG 2518
Db 2319 GCGAGCCACTGCCCCCTTTGTCAAAATGTGACATGAAGCAGGTATCCAGCATGAAATG 2378
Qy 2519 TAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAACAAA 2578
Db 2379 TAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAACAAA 2438
Qy 2579 AAAACAAAACAAAACACATTTCTAAATCTAGAGATACTTTTAAATCTTCAATTTTAG 2638
Db 2439 AAAACAAAACAAAACACATTTCTAAATCTAGAGATACTTTTAAATCTTCAATTTCTCAT 2493

RESULT 6
AAI58169
ID AAI58169 standard; cDNA, 2905 BP.
XX AC
XX AC AAI58169;
XX DT 22-OCT-2001 (first entry)
XX DT
XX DE Human polynucleotide SEQ ID NO 372.
XX DE
XX DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW KW Peripheral nervous system; neuropathy; central nervous system; CNS;
KW KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW KW leukaemia; ss.
XX KW
XX KW Homo sapiens.
XX KW
XX KW WO200153312-A1.
XX PN
XX PD 26-JUL-2001.
XX PD
XX PF 26-DEC-2000; 2000WO-US34263.
XX PF
XX PF 21-JAN-2000; 2000US-0488725.
XX PR
XX PR 25-APR-2000; 2000US-0552117.
XX PR
XX PR 09-JUL-2000; 2000US-0598042.
XX PR
XX PR 19-JUL-2000; 2000US-0620312.
XX PR
XX PR 03-AUG-2000; 2000US-0653450.
XX PR
XX PR 14-SEP-2000; 2000US-0662191.
XX PR
XX PR 19-OCT-2000; 2000US-0693036.
XX PR
XX PR 29-NOV-2000; 2000US-0727344.
XX PR
XX PR (HYSE-) HYSEQ INC.
XX PA
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM39013.
XX DR
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries .
XX PT
XX PS Claim 1; SEQ ID NO 372; 10078pp; English.
XX PS
XX PS The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX XX
XX SQ Sequence 2905 BP; 965 A; 617 C; 691 G; 632 T; 0 other;
Query Match 75.5%; Score 2797.6; DB 22; Length 2905;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2842; Conservative 0; Mismatches 24; Indels 10; Gaps 2
OV 4 CGCAGGACGAGTAGGTGTTAGCAGCTGTGTCGCGCAGGTGCGCTAGGTAGGCGCGGG 63

4 CGCAGGACGAGTAGGTCTTAGCAGCTTGGTCGCGACAGGTGGCGCTAGGTAGAGCGCGGG 63

Db 40 CCGGGTCGACGATTTTCGTAGCAGCTTGGTCGCGACACAGGTGCGCTAGGTAGAGCGCGGG 99
QY 64 ACCTGTGACAGGCGCTGGTAGCAGCGGACAGAGGAAGGGGCTTTTAGCCAGGTAATTTTCAGT 123
Db 100 ACCTGTGACAGGCGTGGTAGCAGCGGACAGAGGAAGGGGCTTTTAGCCAGGTAATTTTCAGT 159
QY 124 GTCTGTAGACAAGATGAATCATCTCCATTTAATAGACGCAATGGACCTCAGTATCATTT 183
Db 160 GTCTGTAGACAAGATGAATCATCTCCATTTAATAGACGCAATGGACCTCAGTATCATTT 219
QY 184 GAGGTAACAGCCAAAGAACTTTCTTGTCAACAAGAACTCATCGCTATTTGGGA 243
Db 220 GAGGTAACAGCCAAAGAACTTTCTTGTCAACAAGAACTCATCGCTATTTGGGA 279
QY 244 AATATTTCCAAATACAGAAAAGCAGCTGAAGAAAACAAACATCGAGAAAGAGAAAGTAA 303
Db 280 AATATTTCCAAATACAGAAAAGCAGCTGAAGAAAACAAACATCGAGAAAGAGAAAGTAA 339
QY 304 CACCGAAAATCTCTCCAGCAGCTTTAGAAAAGGGACCCCTGACTGTGTTTAAAGAAAGTGA 363
Db 340 CACCGAAAATCTCTCCAGCAGCTTTAGAAAAGGGACCCCTGACTGTGTTTAAAGAAAGTGA 399
QY 364 GGAGAACCCAGGGGTGGGACGAGTCTCACACAGACTCTCTACGGAAACAGACGACTGA 423
Db 400 GGAGAACCCAGGGGTGGGACGAGTCTCACACAGACTCTCTACGGAAACAGACGACTGA 459
QY 424 GATTAGSCACAGACAGACCATCCTCCTGCTGAAGTGACAAGCCACGCTGCTTTCTGGAGC 483
Db 460 GATTAGSCACAGACAGACCATCCTCCTGCTGAAGTGACAAGCCACGCTGCTTTCTGGAGC 519
QY 484 CAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTACCTCCTGAAGC 543
Db 520 CAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTACCTCCTGAAGC 579
QY 544 CCTCGTTTCAGGTCGATATCCCGACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAAC 603
Db 580 CCTCGTTTCAGGTCGATATCCCGACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAAC 639
QY 604 AGAAAGTAAAAAATGGAAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGA 663
Db 640 AGAAAGTAAAAAATGGAAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGA 699
QY 664 AATCAGTGAACAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACAG 723
Db 700 AATCAGTGAACAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACAG 759
QY 724 GCTTAAGATGATGTTGAGAAAAGGTGNAACCACTCAAACTAAGATTTCTCGGGCCCCAAG 783
Db 760 GCTTAAGATGATGTTGAGAAAAGGTGNAACCACTCAAACTAAGATTTCTCGGGCCCCAAG 819
QY 784 CCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAAT 843
Db 820 CCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAAT 879
QY 844 AGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAA 903
Db 880 AGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAA 939
QY 904 TCTGGAACTTCCAGCGCTCTCAGAAAACCTCTATAAGAGATCGAATGGCCAAAGTACCAGGC 963
Db 940 TCTGGAACTTCCAGCGCTCTCAGAAAACCTCTATAAGAGATCGAATGGCCAAAGTACCAGGC 999
QY 964 AGCTGTGTCCAAACAAGCAGCTCAACCACTATACAAATGAGCTCAAAAGCCAGTGGTG 1023
Db 1000 AGCTGTGTCCAAACAAGCAGCTCAACCACTATACAAATGAGCTCAAAAGCCAGTGGTG 1059
QY 1024 CGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTG 1083
Db 1060 CGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTG 1119
QY 1084 CATCAACCATCAGGAAGGGGAAAAGATTTCTGCAAAATGAGAAATGAGCTTGGCAGTCCGTTT 1143
Db 1120 CATCAACCATCAGGAAGGGGAAAAGATTTCTGCAAAATGAGAAATGAGCTTGGCAGTCCGTTT 1179

QY 1144 CACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGT 1203
Db 1180 CACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGT 1239
QY 1204 CCATCCCAAGCCACTAAGTTCAGAGATTTCAGAGCTCCAGTCTTTCTGAAAGTTCCTCTCC 1263
Db 1240 CCATCCCAAGCCACTAAGTTCAGAGATTTCAGAGCTCCAGTCTTTCTGAAAGTTCCTCTCC 1299
QY 1264 CAAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAAGAGAC 1323
Db 1300 CAAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAAGAGAC 1359
QY 1324 AGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAGAGTGTTCACATCAGCTGCTTCG 1383
Db 1360 AGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAGAGTGTTCACATCAGCTGCTTCG 1419
QY 1384 TTGCTCTTATTGCAACACAAACTCAGTCTAGGACATATGCATCTTTTACATGGAGAAAT 1443
Db 1420 TTGCTCTTATTGCAACACAAACTCAGTCTAGGACATATGCATCTTTTACATGGAGAAAT 1479
QY 1444 CTATTTGTAAGGCTCACTTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGCTT 1503
Db 1480 CTATTTGTAAGGCTCACTTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGCTT 1539
QY 1504 TGGGCACAGACACACAAAGGATCTATGGGCAAGCAAAAATGAAAAACGAAGAGATTTTGA 1563
Db 1540 TGGGCACAGACACACAAAGGATCTATGGGCAAGCAAAAATGAAAAACGAAGAGATTTTGA 1599
QY 1564 GAGACCCAGCCAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCGGGGTAGAAGATGC 1623
Db 1600 GAGACCCAGCCAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCGGGGTAGAAGATGC 1659
QY 1624 CCCTATTCTTAAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCTCTCTCTCAGCA 1683
Db 1660 CCCTATTCTTAAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCTCTCTCTCAGCA 1719
QY 1684 GGAGAAAGGAACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTCGCCACCCCCAC 1743
Db 1720 GGAGAAAGGAACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTCGCCACCCCCAC 1779
QY 1744 TGAACCTTGGAAAGTTTCAGGAAGTGCCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAAATG 1803
Db 1780 TGAACCTTGGAAAGTTTCAGGAAGTGCCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAAATG 1839
QY 1804 GCCTCCTGAAGACGAAATCAGAACGCCCGAAGTTCCTCAGGATGTCGATCTAGATCTGAA 1863
Db 1840 GCCTCCTGAAGACGAAATCAGAACGCCCGAAGTTCCTCAGGATGTCGATCTAGATCTGAA 1899
QY 1864 GAAAGTGAAGACGATCTTCTTCACTGAAGGAAGAGCCGCCATTCACCTGTAGCAGCTTC 1923
Db 1900 GAAAGTGAAGACGATCTTCTTCACTGAAGGAAGAGCCGCCATTCACCTGTAGCAGCTTC 1959
QY 1924 ATTTTCAAAGCACCTCTGTCAAGAGCCCCAAAACTGTGTCCCCCACCCTATCAGGAAAAAGGCTG 1983
Db 1960 ATTTTCAAAGCACCTCTGTCAAGAGCCCCAAAACTGTGTCCCCCACCCTATCAGGAAAAAGGCTG 2019
QY 1984 GAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTCAGAGAAAGGAAACAAATG 2043
Db 2020 GAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTCAGAGAAAGGAAACAAATG 2079
QY 2044 GGAAAAATGCCAAGGCTTCTTAAGAGAAATGGGAAATGTGGGAAAAACAACTGGCAAAAACAA 2103
Db 2080 GGAAAAATGCCAAGGCTTCTTAAGAGAAATGGGAAATGTGGGAAAAACAACTGGCAAAAACAA 2139
QY 2104 AGAATCTTAAAGAGAGACAGGGAAGAGAGTAAGGAAGTTCATAGTTTGGAGATGGAGAA 2163
Db 2140 AGAATCTTAAAGAGAGACAGGGAAGAGAGTAAGGAAGTTCATAGTTTGGAGATGGAGAA 2199
QY 2164 TGAGAAATCTTTAGAAAAATGGTCAGAGCTCCGATGAAGATGATAACAGCTTCCTCTCAACA 2223
Db 2200 TGAGAAATCTTTAGAAAAATGGTCAGAGCTCCGATGAAGATGATAACAGCTTCCTCTCAACA 2259

1 Jan 7 10:31:29 2004

2224 ACAATCTCCACAAGAACCAAGTCTCTGAATGGTCGAGTTTGTAGACAAACACTTTGC 2283
2260 ACAATCTCCACAAGAACCAAGTCTCTGAATGGTCGAGTTTGTAGACAAACACTTTGC 2319
2284 TGAAGAAATTCACCTACTCAGAAATCCAGGATGCGAACTCTGGAGGGAAGT 2343
2320 TGAAGAAATTCACCTACTCAGAAATCCAGGATGCGAACTCTGGAGGGAAGT 2379
2344 GGTCAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATATGATGAGATCA 2403
2380 GGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATATGATGAGATCA 2439
2404 GGTCAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATATGATGAGATCA 2463
2440 GGTCAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATATGATGAGATCA 2499
2464 CCACCTGCCCTTTGTCAAAATGTCATGATGATGATGATGATGATGATGATGATGAT 2523
2500 CCACCTGCCCTTTGTCAAAATGTCATGATGATGATGATGATGATGATGATGATGAT 2559
2524 TACTTGGAGTAACTTTGGAAGAAATTCCTTCTTAAATCAAAACAAACAAAC 2583
2560 TACTTGGAGTAACTTTGGAAGAAATTCCTTCTTAAATCAAAACAAACAAAC 2612
2584 ACAAAACACATTTCTAAATCTAGAGATTAATCTTAAATCTTAAATCTTAAATCT 2643
2613 - - - - - 2670
2644 ATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2703
2671 ATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2729
2704 CAGATTTCTACTGATTTCCCAAAAGGCAATATTAAGGTAGATGATGATGATGATGAT 2763
2730 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 2789
2764 GTTACAGACTATTTTGGAAATTAGAGAAATCAGAGAGAAATTTAGGGGCTTTAAACATTAC 2823
2790 GTTACAGACTATTTTGGAAATTAGAGAAATCAGAGAGAAATTTAGGGGCTTTAAACATTAC 2849
2824 GACTGAATGCACCTTTAGTATTAAGGGCAGATTTGTATATTTTAAATGAATACCA 2879
2850 GACTGAATGCACCTTTAGTATTAAGGGCAGATTTGTATATTTTAAATGAATACCA 2905

SULT 7

AA572746 standard; cDNA; 3465 BP.

AA572746;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #8550.

Human; chromosome mapping; gene mapping; gene therapy; forensic;

food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.
XX P-PSDB; ABG08559.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity

Claim 1; SEQ ID No 8550; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX diagnostics, forensics, gene mapping, identification of mutations
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX a food supplement. (II) and its binding partners are useful in medical
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity
XX the polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3465 BP; 1112 A; 706 C; 773 G; 874 T; 0 other;

Query Match 75.3%; Score 2789.2; DB 23; Length 3465;

Best Local Similarity 94.3%; Pred. No. 0;

Matches 3266; Conservative 0; Mismatches 123; Indels 73; Gaps 33;

302 AACACCGGAAATCTCCAGCAGCTTTAGAAAGGACCCCTGACTGTGTAAAGAGAAG 361

4 AACACCGGAAATCTCCAGCAGCTTTAGAAAGGACCCCTGACTGTGTAAAGAGAAG 63

362 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTACGGACAGCAGCACT 421

64 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTACGGACAGCAGCACT 123

422 GAGATTAGGCACAGAGCAGACCATCTCTGCTGAAGTGACAGCCACGGCTCTTCTGGA 481

124 GAGATTAGGCACAGAGCAGACCATCTCTGCTGAAGTGACAGCCACGGCTCTTCTGGA 183

482 GCCAAAGCTGACCAAGAGAGAAACAAATCCACCCGATCTAGACTCAGGTCACTCTCTGAA 541

184 GCCAAAGCTGACCAAGAGAGAAACAAATCCACCCGATCTAGACTCAGGTCACTCTCTGAA 243

542 GCCCTCGTTTCAAGGTGATATCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCA 601

244 GCCCTCGTTTCAAGGTGATATCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCA 303

602 ACAGAAAGTAAAAAATGGAATTTCTTAGAGAAATCCAGGATCAAGTAGAAAAATCA 661

304 ACAGAAAGTAAAAAATGGAATTTCTTAGAGAAATCCAGGATCAAGTAGAAAAATCA 363

662 GAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAAC 721

364 GAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAAC 423

722 AGGCTTAAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACTAAGATTTCCGGGCCCAA 781

424 AGGCTTAAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACTAAGATTTCCGGGCCCAA 483

782 AGCCGAAAGTGAAGTGAAGGAGATCTCTGAAACACAGCTATTTCTCTAGATGACCTGAA 841

484 AGCCGAAAGTGAAGTGAAGGAGATCTCTGAAACACAGCTATTTCTCTAGATGACCTGAA 543

QY 842 ATAGCCCCAGGTCAGTTGTCTCTCTACA-TTTGACTCGGAGAAATGAGAGTAGACG 900
DB 544 ATAGCCCCAGGTCAGTTGTCTCTCTCTCATTTTGTCTCGGAGAAATGAGAGTAGACG 603
QY 901 AATCTGGAAGCTTCCAGCCCTCTCAGAACTCTATAAAGATCGAATGGCCAAAGTACCA 960
DB 604 AAATCTGGAAGCTTCCAGCCCTCTCAGAACTCTATAAAGATCGAATGGCCAAAGTACCA 663
QY 961 GGCAGCTGTGTCCTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGG 1020
DB 664 GGCAGCTGTGTCCTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAATGCCAGCGG 723
QY 1021 TGGCGAAATCAAAATTCATATAAATGGAGCAAAAGAGAGAAATGTGCCCCAGTCTGTAGGT 1080
DB 724 TGGCGAAATCAAAATTCATATAAATGGAGCAAAAGAGAGAAATGTGCCCCAGTCTGTAGGT 783
QY 1081 CTGCATACCCCA-TCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATGAGCTGGCAGTCC 1139
DB 784 CTGCATACCCCAATTCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATGAGCTGGCAGTCC 843
QY 1140 GTTCCACCCCTGCCGAAGTAGTCTCC---GTGACTCCAGGTTAAGAGTAGGTTCAAC 1196
DB 844 GTTCCACCCCTGCCGAAGTAGTCTCCAGGTGACTCCAGGTAAAGAGTAGGTTCAAC 903
QY 1197 AGCCTGTGCATCCCAAGGCACCTAAGTCCAGATTCAGAGCCTCAGTCTTTCTGAAAGTT 1256
DB 904 AGCCTGTGCATCCCAAGGCACCTAAGTCCAGATTCAGAGCCTCAGTCTTTCTGAAAGTT 963
QY 1257 CTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTCGCTGGAAATGTC 1316
DB 964 CTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTCGCTGGAAATGTC 1023
QY 1317 AGAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCAATCACTAGCT 1376
DB 1024 AGAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCAATCACTAGCT 1083
QY 1377 GCTTCCGTTGCTCCTATTGCAACAAC-AAACTCAGTCTAGGAACATATGCATCTTTACAT 1435
DB 1084 GCTTCCGTTGCTCCTATTGCAACAACAATACTCAGTCTAGGAACATATGCATCTTTACAT 1143
QY 1436 GGAAGAACTTATGTAAAGCCTCAGTCTCAATCAACTCTTTTAAATCT--AAGGCACTATG 1493
DB 1144 GGAAGAACTTATGTAAAGCCTCAGTCTCAATCAACTCTTTTAAATCTTAAGGGCAAACTATG 1203
QY 1494 ATGAGGCTTTGGGCACAGACCACACAGGATCT--ATGGGCAAGCAAAATGAAAC-G 1550
DB 1204 ATGAGGCTTTGGGCACAGACCACACAGGATCTTTATGGGCAAGCAAAATGAAACCGG 1263
QY 1551 AAGAGATTTTGAGAGACACAG-CCAGCTTGCAAAATGCAAGGAGACCCCTCACAGCCCA 1609
DB 1264 AGGAGATTTTGAGAGACACAGCCCGAGCTTGCAAAATGCAAGGAGACCCCTCACAGCCCA 1323
QY 1610 GGGGTAGAAGATGCCCTTATGCTAAGGTGGGTGCTCGCTGCAAGTATGGAAGCCCAAG 1669
DB 1324 GGGGTAGAAGATGCCCTTATGCTAAGGTGGGTGCTCGCTGCAAGTATGGAAGCCCAAG 1383
QY 1670 GCCTCTCTCAGCAGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCC 1729
DB 1384 GCCTCTCTCAGCAGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCC 1443
QY 1730 TGGCCACCCCACTCAACTTGGAACTTCAGGAAGTCCCTGAGGAAGGGATCAAAATG 1789
DB 1444 TGGCCACCCCACTCAACTTGGAACTTCAGGAAGTCCCTGAGGAAGGGATCAAAATG 1503
QY 1790 TCAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCCTCAGGATGTC 1849
DB 1504 TCAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCCTCAGGATGTC 1563
QY 1850 GATCTAGATCTGAAGAGCTTAAGACGATCTTTCTACTGAAGAAAGAGCCGCCCAATTC 1909
DB 1564 GATCTAGATCTGAAGAGCTTAAGACGATCTTTCTACTGAAGAAAGAGCCGCCCAATTC 1623

QY 1910 ACTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAGAGCCCAAAACCTGTGTCCCAAC- 1968
DB 1624 ACTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAGAGCCCAAAACCTGTGTCCCAACCT 1683
QY 1969 TATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGGAAAGATTGC 2028
DB 1684 TATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGGAAAGATTGC 1743
QY 2029 AGAAAGGAAACAAGTGGAAATGCCAAGCTTCTAAGAAAGATGGGAATGTGGGAAAAAC 2088
DB 1744 AGAAAGGAAACAAGTGGAAATGCCAAGCTTCTAAGAAAGATGGGAATGTGGGAAAAAC 1803
QY 2089 AACCTGGCAAAACAAAGAAATCTAAAGGAGAGACAGGAAAGAGAAAGTAAGG--AAGGTCTAT 2146
DB 1804 AACCTGGCAAAACAAAGAAATCTAAGGAGAGACAGGAAAGAGAAAGTAAGGAAAGTCCAT 1863
QY 2147 AGTTTGAGATGGAGAAATGAGAAATCTT-----GTAGAAATGGTGCAGACTCCGATGAAG 2201
DB 1864 AGTTTGAGATGGAGAAATGAGAAATCTTGTAGGAAATGGTGCAGACTCCGATGAAG 1923
QY 2202 ATGATAACAGCTCTCTCAAACAACAA-----TCTCCAAAGAACCCCAAGTC----TCTGA 2252
DB 1924 ATGATAACAGCTCTCTCAAACAACAACTTCTCCAAAGAACCCCAAGTCTTCTTGAAT 1983
QY 2253 ATTGGTCGAGCTTTGTAGACAACACAC---TTTGTCTGAAGAAATTCACCTACT-CAGAAATCAG 2308
DB 1984 TTGTCGAGTCTTTGTAGACAACACACCTTTTGTCTGAAGAAATTCACCTACTCCAGAAATCAG 2043
QY 2309 AAATCCCAGGATGTGGAACCTCTGGGAGGAGAAAGTGGTCAAAGAGTCTCTCTGTGGAGAA 2368
DB 2044 AAATCCCAGGATGTGGAACCTCTGGGAGGAGAAAGTGGTCAAAGAGTCTCTCTGTGGAGAA 2103
QY 2369 CAGATAAGAGAAATCGTATTTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATG 2428
DB 2104 CAGATAAGAGAAATCGTATTTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATG 2163
QY 2429 ATGCT--GGGCTTTAAATTCATGTTAGTGTAGCGAGCCTGCGCTTTGTCAAAATGTGA 2487
DB 2164 ATGCTGGGCTTTAAATTCATGTTAGTGTAGCGAGCCTGCGCTTTGTCAAAATGTGA 2223
QY 2488 TGCATATAACGAGTATCCGAGATGAAATGTAATTTACTTTGGAAGTAACCTTTGGAAAAAG 2547
DB 2224 TGCATATAACGAGTATCCGAGATGAAATGTAATTTACTTTGGAAGTAACCTTTGGAAAAAG 2283
QY 2548 AATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAA--ACACATTTCTTAAATAC 2605
DB 2284 AATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAACACACATTTCTTAAATAC 2343
QY 2606 TAGAGATAAATTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGATAGTCTGTAA 2665
DB 2344 TAGAGATAAATTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGATAGTCTGTAA 2403
QY 2666 GCCTGTAACTGGGAAATATTCACCTGATATAGCCAGATTCCTACTGTATTCCTCAA 2724
DB 2404 GCCTGTAACTGGGAAATATTCACCTGATATAGCCAGATTCCTACTGTATTCCTCAA 2463
QY 2725 AAGCAATATTAAGSTAGATA-GATGATTAG-TAGTATTTTGTACACACTATTTTGGAA 2782
DB 2464 AAGCAATATTAAGSTAGGAGGATGATTAGTATATTTTGTACACACTATTTTGGAA 2523
QY 2783 TTAGAGAACATACAGAGGAATTTAGGGGCTTTAAACATTAACGACT-GAATGCACTTTAGT 2841
DB 2524 TTAGAGAACATACAGAGGAATTTAGGGGCTTTAAACATTAACGACTGGAATGCACTTTAGT 2583
QY 2842 ATAAAGGCGACAGTTTGTATTAATTTT---AAATGAATACCAATTTAATTTTGTAGTATTT 2898
DB 2584 ATAAAGGCGACAGTTTGGTATATTTTAAATGGAATACCAATTTAATTTTGTAGTATTT 2643
QY 2899 ACCTGTTAAGAGATTTTAGTC-TTTAAATTTTGTAGTTAATTTTCTGCTGTGATAT 2957
DB 2644 ACCTGTTAAGAGATTTTAGTCTTTTAAATTTTGTAGTTAATTTTCTGCTGTGATAT 2703
QY 2958 ATATGAGGAATTTTACTACTTTTATGCTCTGCTCTC-TAAACTACATCCT-GAACTCGACGT 3015

QY 1361 GTGTTTACATCAGCTGCTTCGGTTTGTCTCTATTGCAACACAACTCAGTCTAGGAACA 1420
Db |||||
QY 420 GTGTTTACATCAGCTGCTTCGGTTTGTCTCTATTGCAACACAACTCAGTCTAGGAACA 479
Db |||||
QY 1421 TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACCTTCAATCAACTCTTTAAATCT 1480
Db |||||
QY 480 TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACCTTCAATCAACTCTTTAAATCT 539
Db |||||
QY 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGAGATCTATGGCGAAGCAAA 1540
Db |||||
QY 540 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGAGATCTATGGCGAAGCAAA 599
Db |||||
QY 1541 AATGAAACAGAGAGATTTTGGAGAGACCACCCAGCTTGCNAATGCAAGGGAGACCCCT 1600
Db |||||
QY 600 AATGAAACAGAGAGATTTTGGAGAGACCACCCAGCTTGCNAATGCAAGGGAGACCCCT 659
Db |||||
QY 1601 CACAGCCAGGGGTAGAGATGCCCCCTATTGCTAAAGTGGGTGCTCGGTGCAAGTATG 1660
Db |||||
QY 660 CACAGCCAGGGGTAGAGATGCCCCCTATTGCTAAAGTGGGTGCTCGGTGCAAGTATG 719
Db |||||
QY 1661 GAAGCCAAAGGCTCTCTCAGCAGGAGAGAAAGACCAAGCCAGCTGAACCAAGAGCTG 1720
Db |||||
QY 720 GAAGCCAAAGGCTCTCTCAGCAGGAGAGAAAGACCAAGCCAGCTGAACCAAGAGCTG 779
Db |||||
QY 1721 AGGATCCGCTGGCCACCCCTCACTGAACCTTGGAAAGTTTCAGAAAGTGCCTTGGAGGAAGG 1780
Db |||||
QY 780 AGGATCCGCTGGCCACCCCTCACTGAACCTTGGAAAGTTTCAGAAAGTGCCTTGGAGGAAGG 839
Db |||||
QY 1781 ATCAAAATGTCAAAAGCCCAATAGCCCTCTTGAAGACGAAATCAGCAAGCCCGAAGTTCCT 1840
Db |||||
QY 840 ATCAAAATGTCAAAAGCCCAATAGCCCTCTTGAAGACGAAATCAGCAAGCCCGAAGTTCCT 899
Db |||||
QY 1841 GAGGATTCGATCTGATGAAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAAGC 1900
Db |||||
QY 900 GAGGATTCGATCTGATGAAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAAGC 959
Db |||||
QY 1901 CGCCCATCTAGTGTAGCAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAACTGTG 1960
Db |||||
QY 960 CGCCCATCTAGTGTAGCAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAACTGTG 1019
Db |||||
QY 1961 TCCCCACCTATCAGGAAGGCTGGAGCATCTCAGACAGAGTCAAGAGTCTGTGGTGGA 2020
Db |||||
QY 1020 TCCCCACCTATCAGGAAGGCTGGAGCATCTCAGACAGAGTCAAGAGTCTGTGGTGGA 1079
Db |||||
QY 2021 AGAGTTGCAGAAAGGAACAAAGTGGAAAATGCCAAGGCTCTTAAGAAGAAATGGGAATGTG 2080
Db |||||
QY 1080 AGAGTTGCAGAAAGGAACAAAGTGGAAAATGCCAAGGCTCTTAAGAAGAAATGGGAATGTG 1139
Db |||||
QY 2081 GGAAAAACAACCTGGCNAACAAAGAAATCTAAAGGAGACAGGGGAAGAGAAAGGAA 2140
Db |||||
QY 1140 GGAAAAACAACCTGGCNAACAAAGAAATCTAAAGGAGACAGGGGAAGAGAAAGGAA 1199
Db |||||
QY 2141 GGTCTAGTTTGGAGATGGAGATGGAATCTTTGTAGAAAATGGTGCAGCTCCCGATGAA 2200
Db |||||
QY 1200 GGTCTAGTTTGGAGATGGAGATGGAATCTTTGTAGAAAATGGTGCAGCTCCCGATGAA 1259
Db |||||
QY 2201 GATGATAACAGCTTCTTCAAAACAACAACTTCCACAAGAACCCCAAGTCTCTGAATTTGGTCG 2260
Db |||||
QY 1260 GATGATAACAGCTTCTTCAAAACAACAACTTCCACAAGAACCCCAAGTCTCTGAATTTGGTCG 1319
Db |||||
QY 2261 AGTTTGTAGACAACACCTTGTCTGAAGAAATTCACCTACTCAGAAATCCAGGAT 2320
Db |||||
QY 1320 AGTTTGTAGACAACACCTTGTCTGAAGAAATTCACCTACTCAGAAATCCAGGAT 1379
Db |||||
QY 2321 GTGGAACTCTGGAGGAGAAAGTGTCAAGAGCTCTCTGTGGAAGAAACAGATAAGAGA 2380
Db |||||
QY 1380 GTGGAACTCTGGAGGAGAAAGTGTCAAGAGCTCTCTGTGGAAGAAACAGATAAGAGA 1439
Db |||||
QY 2381 AATCGGTATTATGATCAGGATGAGGATGAGAGTGCACAAATTCGATGATGCTGGGCCCT 2440
Db |||||
QY 1440 AATCGGTATTATGATCAGGATGAGGATGAGAGTGCACAAATTCGATGATGCTGGGCCCT 1499
Db |||||

QY 2441 AAATTCATGTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
Db |||||
QY 1500 AAATTCATGTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG 1559
Db |||||
QY 2501 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTACCTTTGGAAA*GAATTCCTCTTAA 2560
Db |||||
QY 1560 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAGAATTCCTCTTAA 1619
Db |||||
QY 2561 AATCAAAAACAAAAACAAAAACAAAAACACATTTCTAAATACTAGAGATAAATTTAC 2620
Db |||||
QY 1620 AATCAAAAACAAAAACAAAAACAAAAACACATTTCTAAATACTAGAGATAAATTTAC 1679
Db |||||
QY 2621 TTAATTCCTTCAATTTTACAGTGTATGATATGATGCTGTAAAGCTTTGTAACTGGGG 2680
Db |||||
QY 1680 TTAATTCCTTCAATTTTACAGTGTATGATGCTGTAAAGCTTTGTAACTGGGG 1739
Db |||||
QY 2681 AATATTCACCTGATAATAGCCAGATCTACTGTATTTCCCAAAAGGCAATATTAAAGT 2740
Db |||||
QY 1740 AATATTCACCTGATAATAGCCAGATCTACTGTATTTCCCAAAAGGCAATATTAAAGT 1799
Db |||||
QY 2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTTAGAACATACAGAG 2800
Db |||||
QY 1800 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTTAGAACATACAGAG 1859
Db |||||
QY 2801 GAATTTAGGGGCTTAAACACTTACGACTGAATGCACCTT*AGTATAAAGGCCACAGTTGTA 2860
Db |||||
QY 1860 GAATTTAGGGGCTTAAACACTTACGACTGAATGCACCTT*AGTATAAAGGCCACAGTTGTA 1919
Db |||||
QY 2861 TATTTTAAATGAATACCAATTAATTTT*TTTAGTATTTTACCTGTTAAGAGATATTATAGT 2920
Db |||||
QY 1920 TATTTTAAATGAATACCAATTAATTTT*TTTAGTATTTTACCTGTTAAGAGATATTATAGT 1979
Db |||||
QY 2921 CTTTAAATTTT*TTTAGTATTTTCTTCTGCTGTATATATAGGAAATTTACTACTTTAT 2980
Db |||||
QY 1980 CTTTAAATTTT*TTTAGTATTTTCTTCTGCTGTATATATAGGAAATTTACTACTTTAT 2039
Db |||||
QY 2981 GTCCTGCTCTCTAAACTTACATCCTGAACCTCGACGCTCTGAGGTATAATACACAGAGCAC 3040
Db |||||
QY 2040 GTCCTGCTCTCTAAACTTACATCCTGAACCTCGACGCTCTGAGGTATAATACACAGAGCAC 2099
Db |||||
QY 3041 TTTT*GAGCAATTGAAAACCAACCTACACTCTTTCGGTGTCTTAGAGAGATCTGTGCT 3100
Db |||||
QY 2100 TTTT*GAGCAATTGAAAACCAACCTACACTCTTTCGGTGTCTTAGAGAGATCTGTGCT 2159
Db |||||
QY 3101 CCCAAAT*AAAGCTTTTGTATCTGCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 3160
Db |||||
QY 2160 CCCAAAT*AAAGCTTTTGTATCTGCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 2219
Db |||||
QY 3161 TCTGTGTATATCTGTGCTTCTCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC 3220
Db |||||
QY 2220 TCTGTGTATATCTGTGCTTCTCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC 2279
Db |||||
QY 3221 GGGATCACTGTCCCCCATCTTCCGTTT*GAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA 3280
Db |||||
QY 2280 GGGATCACTGTCCCCCATCTTCCGTTT*GAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA 2339
Db |||||
QY 3281 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 3340
Db |||||
QY 2340 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 2399
Db |||||
QY 3341 GTCCCTTTT*GAGACACTAAATTTT*AAATACTTACTAGCTCTGAAATATATGATTTTT 3400
Db |||||
QY 2400 GTCCCTTTT*GAGACACTAAATTTT*AAATACTTACTAGCTCTGAAATATATGATTTTT 2459
Db |||||
QY 3401 ATCAGATATTCTCAGGGTGAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTT 3460
Db |||||
QY 2460 ATCAGATATTCTCAGGGTGAATTTAAACCAACTATAGGCCCTTTTCTTGGGATGATTTT 2519
Db |||||
QY 3461 CTAGCTTAAAGGTTTGGGACATTATAAACCTT*GAGTACATTTGTTGTACACAGTTGATAT 3520
Db |||||
QY 2520 CTAGCTTAAAGGTTTGGGACATTATAAACCTT*GAGTACATTTGTTGTACACAGTTGATAT 2579
Db |||||
QY 3521 TCCAAAT*GTATGATGGAGGGAGAGGTGTCTTAAAGCTGTAGGCTTTTCTTGTACTGC 3580
Db |||||

2580 TCCAAATTGATGATGGAGGAGAGGTGCTTAAGCTGAGGCTTTCTTTGTACTGC 2639
3581 ATTTATAGAGATTAGCTTTAATATTTTATAGAGATGTAAGCAATTCGTCTTTCTAGTC 3640
2640 ATTTATAGAGATTAGCTTTAATATTTTATAGAGATGTAAGCAATTCGTCTTTCTAGTC 2699
3641 TTACCTAGTCTGAACATTTTATTCATTAAGAGATTTTAAATTAATAATTG 3690
2700 TTACCTAGTCTGAACATTTTATTCATTAAGAGATTTTAAATTAATAATTG 2749
LT 9
3566
AAZ33566 standard; cDNA; 2783 BP.
AAZ33566;
08-DEC-1999 (first entry)
Human breast tumour-associated EST 26.
Expressed sequence tag; EST; human; breast; cancer; cytostatic;
medicaments; gene therapy; treatment; fat metabolism; ss.
Homo sapiens.
DE19813835-A1.
23-SEP-1999. 98DE-1013835.
20-MAR-1998; 98DE-1013835.
20-MAR-1998; 98DE-1013835.
(META-) METAGEN GES GENOMFORSCHUNG MBH.
Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
WPI: 1999-528979/45.
P-PSDB; AAY48487.
Human nucleic acid sequences and protein products from normal breast
tissue, useful for breast cancer therapy
Claim 1a; 113-114; 206pp; German.
This invention describes novel human nucleic acid sequences from normal
breast tissue which have cytotostatic activity. The nucleic acid sequences
can be used to produce and isolate full-length gene sequences. They can
be used to express proteins, which can be used as tools to find an
activity against breast cancer. The sequences can be used in sense or
antisense form. They are especially useful for medicaments for gene
therapy to treat breast cancer and for treating illnesses associated
with fat metabolism. AAZ33541-233610 represent expressed sequence tags
described in the method of the invention.
Sequence 2783 BP; 901 A; 527 C; 608 G; 747 T; 0 other;
Query Match 72.8%; Score 2696.6; DB 20; Length 2783;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1001 AATGAGCTGAAGAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGCAAT 1060
40 AATGAGCTGAAGAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGCAAT 99
1061 GTGCCCCCAGGCTCTGAGGTCTGCAATCCCATCAGGAAGGGAAAGATTCTCGCAAT 1120
100 GTGCCCCCAGGCTCTGAGGTCTGCAATCCCATCAGGAAGGGAAAGATTCTCGCAAT 159
1121 GAGATAGCTGGCAGTCCGTTCCACCTGCGGAGATGACTCCCGTGACTCCAGGTT 1180

Db 160 GAGAAATAGCTGGCAGTCCGTTCCACCTCCGGAAGATGACTCCCGTGACTCCAGGTT 219
QY 1181 AAGAGTGAAGTTTCAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCC 1240
Db 220 AAGAGTGAAGTTTCAACAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCC 279
QY 1241 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAAGTTTTCAGGCACCTGCAAGAGAG 1300
Db 280 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAAGTTTTCAGGCACCTGCAAGAGAG 339
QY 1301 ACCTGGTGGMATGTCAGAAAGCAGTCTATCCAATGGAGCGTCTCTTGGCCACCAAGCAG 1360
Db 340 ACCTGGTGGMATGTCAGAAAGCAGTCTATCCAATGGAGCGTCTCTTGGCCACCAAGCAG 399
QY 1361 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTTGCAACAACAACTCAGTCTAGGAACA 1420
Db 400 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTTGCAACAACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
Db 460 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1540
Db 520 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 579
QY 1541 AATGAAAACGAAGAGATTTTGGAGAGACCCAGCTTGCAAAATGGAAGGAGAGACCCCT 1600
Db 580 AATGAAAACGAAGAGATTTTGGAGAGACCCAGCTTGCAAAATGGAAGGAGAGACCCCT 639
QY 1601 CACAGCCCAAGGCTTGAAGATGCCCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATG 1660
Db 640 CACAGCCCAAGGCTTGAAGATGCCCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATG 699
QY 1661 GAAGCCCAAGGCTTCTCTCAGCAGGAGGAAGAGACCAAGCCAGCTGAAACCAAGAGCTG 1720
Db 700 GAAGCCCAAGGCTTCTCTCAGCAGGAGGAAGAGACCAAGCCAGCTGAAACCAAGAGCTG 759
QY 1721 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTCCCTTGAGGAGGG 1780
Db 760 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTCCCTTGAGGAGGG 819
QY 1781 ATCAAAATGTCAAAGCCCAAAATGGCTCTCTGAAGAGCAATCAGCAAGCCCAAGTTCTCT 1840
Db 820 ATCAAAATGTCAAAGCCCAAAATGGCTCTCTGAAGAGCAATCAGCAAGCCCAAGTTCTCT 879
QY 1841 GAGGATCTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1900
Db 980 GAGGATCTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 939
QY 1901 CGCCATTCATCTGAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960
Db 999 CGCCATTCATCTGAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 999
QY 1961 TCCCCACCTATCAGGAAGGCTGGAGCATGTGAGGAGAGTGAAGTCTGTGGGTGGA 2020
Db 1000 TCCCCACCTATCAGGAAGGCTGGAGCATGTGAGGAGAGTGAAGTCTGTGGGTGGA 1059
QY 2021 AGAGTTGCAAGAAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTG 2080
Db 1060 AGAGTTGCAAGAAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTG 1119
QY 2081 GGAAAGAAACCTGCAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTG 2140
Db 1120 GGAAAGAAACCTGCAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAAATGGGAATGTG 1179
QY 2141 GGTCATAGTTTGGAGATGGAGATGAGATCTTTAGAAAATGGTGCAGACTCCGATGAA 2200
Db 1180 GGTCATAGTTTGGAGATGGAGATGAGATCTTTAGAAAATGGTGCAGACTCCGATGAA 1239
QY 2201 GATGATAACAGCTTCTCCAAACAACATCTCCCAAGAGCCCAAGTCTCTGAATTTGGTCG 2260
Db 1240 GATGATAACAGCTTCTCCAAACAACATCTCCCAAGAGCCCAAGTCTCTGAATTTGGTCG 1299

The polynucleotide sequences given in AAC59449 to AAC59497 encode the human secreted proteins given in AAB34092 to AAB34140. AAB34141 to AAB34216 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissue and cells the genes are expressed in. Examples of activities include: antiarthritic; immunosuppressive; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebrotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; and vulnerary. The polynucleotides and polypeptides can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders e.g. neoplasms or cancer of the breast or liver, cardiovascular disorders, infections caused by disorders, angiogenesis, nervous system disorders. The polypeptides can bacteria, viruses and fungi and ocular disorders. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used in the exemplification of the present invention.

Sequence 2267 BP; 741 A; 416 C; 490 G; 617 T; 3 other;

Query Match	Best Local Similarity	99.9%; Pred. No. 0;	Mismatches	0;	Indels	0;	Gaps	0;
Matches 2207;	Conservative							
1496	GAAGGCTTTGGGCACACACACACAGGATCTATGGCAGCAAAATGAAACAAAGAG	1555						
32	GAAGGCTTTGGGCACACACACACAGGATCTATGGCAGCAAAATGAAACAAAGAG	91						
1556	ATTTTGAGAGACAGCCAGCTTGCAATGCAAGGAGACCCCTCACAGCCAGGGGTA	1615						
92	ATTTTGAGAGACAGCCAGCTTGCAATGCAAGGAGACCCCTCACAGCCAGGGGTA	151						
1616	GAAGATCCCTATTGTAAGGTGGGTGCTGCTGGCTGCAAGTATGAGCCAGGCTCC	1675						
152	GAAGATCCCTATTGTAAGGTGGGTGCTGCTGGCTGCAAGTATGAGCCAGGCTCC	211						
1676	TCTCAGCAGGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCCA	1735						
212	TCTCAGCAGGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCCA	271						
1736	CCCCCCTGAACTTTGGAAGTTCAGGAAGTGCCTTGGAGGAGGAGATCAAAATGTC	1795						
272	CCCCCCTGAACTTTGGAAGTTCAGGAAGTGCCTTGGAGGAGGAGATCAAAATGTC	331						
1796	CCCAATGGCCTCTGAGACCGAAATCAGACGCCGGAAGTTCCTGAGGATGTCGATCTA	391						
332	CCCAATGGCCTCTGAGACCGAAATCAGACGCCGGAAGTTCCTGAGGATGTCGATCTA	1915						
1856	GATCTGAAGAGCTAGACGATCTTCTTCACTGAAGAGAGCCGCCCTTCACTGTA	451						
392	GATCTGAAGAGCTAGACGATCTTCTTCACTGAAGAGAGCCGCCCTTCACTGTA	1975						
1916	GCAGCTTCATTTCAAGACCTCTGTCAGAGCCGCAAACTGTGTCCCACTATCAGG	511						
452	GCAGCTTCATTTCAAGACCTCTGTCAGAGCCGCAAACTGTGTCCCACTATCAGG	2035						
1976	AAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGAAGAGTTGCAGAAAGG	571						
512	AAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGAAGAGTTGCAGAAAGG	2095						
2036	AAACAAGTGGAATGCAAGGCTTCTAAGAAGATGGGAATGGGAATAAACAACCTGG	631						
572	AAACAAGTGGAATGCAAGGCTTCTAAGAAGATGGGAATGGGAATAAACAACCTGG	2155						
2096	CAAAACAAAGAAATCTAAAGGAGACACAGGAGAGAGTAGGAGAGCTCATAGTTTGGAG							

632	CAAAACAAAGAAATCTAAAGGAGAGACAGGAGAGAAAGTAAAGAAAGTCAATAGTTTGGAG	691
2156	ATGGAGAAATGAGAAATCTTTAGAAAATGGTGCAGATCCGATGAAGATGATAACAGCTTC	2215
692	ATGGAGAAATGAGAAATCTTTAGAAAATGGTGCAGATCCGATGAAGATGATAACAGCTTC	751
2216	CTCAAAACAAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACAAC	2275
752	CTCAAAACAAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACAAC	811
2276	ACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAACTCTGGAG	2335
812	ACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAACTCTGGAG	871
2336	GGAGAAATGGTCAAAAGAGCTCTCTGTGAAAGAACAGATAAAGAGAAATCGGTATTATCAT	2395
872	GGAGAAATGGTCAAAAGAGCTCTCTGTGAAAGAACAGATAAAGAGAAATCGGTATTATCAT	931
2396	GAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGGCTTTAAATTTTATGTTAGTG	2455
932	GAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGGCTTTAAATTTTATGTTAGTG	991
2456	TTAGCGAGCAGCTGCCCTTTGTCAAAATGTGATGCACATGAAGCAGGTATCCAGCATGAA	1051
992	TTAGCGAGCAGCTGCCCTTTGTCAAAATGTGATGCACATGAAGCAGGTATCCAGCATGAA	1111
2516	ATGTAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAC	1171
1052	ATGTAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAC	1231
2576	AAAAAACACAAAAAACACATCTTAATCTAGAGATTAACCTTTTAAATTTTCAATTT	1291
1112	AAAAAACACAAAAAACACATCTTAATCTAGAGATTAACCTTTTAAATTTTCAATTT	1351
2636	TAGCAGTGAATGATGATAGTCTGTAGGCTTTGAAGCTTTGAAGCTTTGAAGCTTTGA	1411
1172	TAGCAGTGAATGATGATAGTCTGTAGGCTTTGAAGCTTTGAAGCTTTGAAGCTTTGA	1471
2696	TAATAGCCAGATCTCTAGTATTTCCCAAAAGGCAATATTAAAGGTAGATGATTAGT	1531
1232	TAATAGCCAGATCTCTAGTATTTCCCAAAAGGCAATATTAAAGGTAGATGATTAGT	1591
2756	AGTATATTGTTACACACTATTTTGAATTTAGAGACATACAGAGGAATTTAGGGGCTTA	1651
1292	AGTATATTGTTACACACTATTTTGAATTTAGAGACATACAGAGGAATTTAGGGGCTTA	1711
2816	AACATTTACGACTGAATGCACTTTAGTATAAGGGGACAGTTTGTATTTTAAATGAAT	1771
1352	AACATTTACGACTGAATGCACTTTAGTATAAGGGGACAGTTTGTATTTTAAATGAAT	1831
2876	ACCAATTTAAATTTTGTAGTATTTTACCTGTTAAGAGATTTTAAATTTTAAATTTTAA	1891
1412	ACCAATTTAAATTTTGTAGTATTTTACCTGTTAAGAGATTTTAAATTTTAAATTTTAA	1951
2936	GTTAATTTTCTGCTGTGATATATAGAGAAATTTACTCTTTATGCTGCTCTCTAA	2011
1472	GTTAATTTTCTGCTGTGATATATAGAGAAATTTACTCTTTATGCTGCTCTCTAA	2071
2996	CTACATCTGAACTCGAGCTCGCTGAGGTATAATACACAGAGCACTTTTGGAGCAATTG	2131
1532	CTACATCTGAACTCGAGCTCGCTGAGGTATAATACACAGAGCACTTTTGGAGCAATTG	2191
3056	AAAAACCAACCTACACTCTTCGGTCTTAGAGATCTGCTGCTCCCAATAAGCTTTT	2251
1592	AAAAACCAACCTACACTCTTCGGTCTTAGAGATCTGCTGCTCCCAATAAGCTTTT	2311
3116	GTATCTGCGAGTGAATTTACTGTACTCCAAATGATTGCTTTTCTGTTGATATCTGT	2371
1652	GTATCTGCGAGTGAATTTACTGTACTCCAAATGATTGCTTTTCTGTTGATATCTGT	2431
3176	GCTTCTCTAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGATCAGCTGCC	2491

1712 GCTTCTCATTAATTAAGTGAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCGCC 1771
3236 CATCTTCCTGTTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGCTCTTA 3295
1772 CATCTTCCTGTTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGCTCTTA 1831
3296 CACCACCTTGAGCTCAGACCTCTAAACCCCTGATTTCCCTTATGATGTCCTCTTTTGAGA 3355
1832 CACCACCTTGAGCTCAGACCTCTAAACCCCTGATTTCCCTTATGATGTCCTCTTTTGAGA 1891
3356 CACTAAATTTTAAATACTACTAGCTCTGAAATATATTTGATTTTATPCACAGATTTCTCA 3415
1892 CACTAAATTTTAAATACTACTAGCTCTGAAATATATTTGATTTTATPCACAGATTTCTCA 1951
3416 GGGTGAATTAACCAACTATAGCCCTTTTCTTTGGGATGATTTCTAGTCTTAAGGTTT 3475
1952 GGGTGAATTAACCAACTATAGCCCTTTTCTTTGGGATGATTTCTAGTCTTAAGGTTT 2011
3476 GGGACATTATAAATCTGAGTACATTTTGTGTACACAGTTGATATTCCTCAATTTGATGGA 3535
2012 GGGACATTATAAATCTGAGTACATTTTGTGTACACAGTTGATATTCCTCAATTTGATGGA 2071
3536 TGGGAGGGAGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTATAGAGATTTA 3595
2072 TGGGAGGGAGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTATAGAGATTTA 2131
3596 GCTTTAATATTTTTPAGAGATGTAAAACAATTCGTCTTTTCTTTAGTCTTACCTAGTCTGAAA 3655
2132 GCTTTAATATTTTTPAGAGATGTAAAACAATTCGTCTTTTCTTTAGTCTTACCTAGTCTGAAA 2191
3656 CATTTTATTCATTAAGATTTTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 3705
2192 CATTTTATTCATTAAGATTTTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 2241

RESULT 11
AAC93483
ID AAC93483 standard; cDNA; 2158 BP.
XX AAC93483;
XX
XX
XX 16-FEB-2001 (first entry)
XX Human secreted protein gene 5 SEQ ID NO:15.
XX
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytosolic; cardiac; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnary; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
XX
XX Homo sapiens.
XX
XX WO200061626-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09066.
XX
XX 09-APR-1999; 99US-0128698.
XX 20-JAN-2000; 2000US-0176926.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-619227/59.
XX P-PSDB; AAB51831.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for

diagnosing, preventing or ameliorating medical conditions and used for
food additives or preservatives -
Claim 1; Page 435; 516pp; English.
CC Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
CC AAB51927 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include immunosuppressive;
CC antirheumatic; antiproliferative; cytosolic; cardiac;
CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
CC virucide; fungicide; ophthalmological; and vulnary. The secreted
CC proteins, polynucleotides, antagonists and agonists may be useful in
CC treating, preventing and/or diagnosing diseases and disorders such as
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular ischaemia,
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrates, vitamins, minerals, cofactors and other nutritional
CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
CC used in the isolation and characterisation of the proteins and
CC polynucleotides of the invention.
XX
XX Sequence 2158 BP; 714 A; 383 C; 458 G; 603 T; 0 other;

Query Match 56.2%; Score 2083.6; DB 21; Length 2158;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2138; Conservative 0; Mismatches 14; Indels 6; Gaps 4;
QY 1553 GAGNTTTTGGAGAGACCCAGCTTTGCAATTTGCAAGGAGAGACCCCTCACAGCCAGGG 1612
DB 1 GAGATTTTGGAGAGACCCAGCTTTGCAATTTGCAAGGAGAGACCCCTCACAGCCAGGG 60
QY 1613 GTAGAAGATGCCCTATTGCTTAAGGTGGTGTCTCTGCTGCTCAAGTATGGAAGCAAGGCC 1672
DB 61 GTAGAAGATGCCCTATTGCTTAAGGTGGTGTCTCTGCTGCTCAAGTATGGAAGCAAGGCC 120
QY 1673 TCCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTGAGGATCGCTGG 1732
DB 121 TCCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTGAGGATCGCTGG 180
QY 1733 CCACCCCTTCACTGAACTTGGAGTTTCAAGTTCAGGAGTGCCTTTGGAGGAGGATCAAAATGTCA 1792
DB 181 CCACCCCTTCACTGAACTTGGAGTTTCAAGTTCAGGAGTGCCTTTGGAGGAGGATCAAAATGTCA 240
QY 1793 AAGCCCAATGGCTCTCTGAAGAGCAAAATCAGCAAGCCGCAAGTTCTCTGAGGATGTCAT 1852
DB 241 AAGCCCAATGGCTCTCTGAAGAGCAAAATCAGCAAGCCGCAAGTTCTCTGAGGATGTCAT 300
QY 1853 CTAGATCTGAAGAAGCTAAGACGATCTTTCTTCACTGAAGGAGAGAGCCGCCCTTCACT 1912
DB 301 CTAGATCTGAAGAAGCTAAGACGATCTTTCTTCACTGAAGGAGAGAGCCGCCCTTCACT 360
QY 1913 GTAGCAGCTTCAATTTCAAAAGCACCTCTGTCAAGAGCCCAAAACCTGTCTCCCACTATC 1972
DB 361 GTAGCAGCTTCAATTTCAAAAGCACCTCTGTCAAGAGCCCAAAACCTGTCTCCCACTATC 420
QY 1973 AGGAAGAGCTGGACATGTCAGAGCAGAGTCAAGAGTCTGTGGGTGGAAGAGTTCGAGAA 2032
DB 421 AGGAAGAGCTGGACATGTCAGAGCAGAGTCAAGAGTCTGTGGGTGGAAGAGTTCGAGAA 480
QY 2033 AGGAAGAGCTGGAGAAATGCCAAGGCTTCTTAAGAGAGATGGGAATGTGGGAGAGAGCAACC 2092
DB 481 AGGAAGAGCTGGAGAAATGCCAAGGCTTCTTAAGAGAGATGGGAATGTGGGAGAGAGCAACC 540

QY 2093 TGGCAAAACAAAGAAATCTTAAGAGAGAGACAGGGAAGAGAAAGTAAAGAGGTCAATAGTTTG 2152
DB 541 TGGCAAAACAAAGAAATCTTAAGAGAGAGACAGGGAAGAGAAAGTAAAGAGGTCAATAGTTTG 600
QY 2153 GAGATGGAGAAATGAGAAATCTTTAGAAATGTTGAGACTCCGATGAAGATGATACAGC 2212
DB 601 GAGATGGAGAAATGAGAAATCTTTAGAAATGTTGAGACTCCGATGAAGATGATACAGC 660
QY 2213 TTCTCAAAACAAATCTCCAC--AAGAACCAGTCTCTGAATGGTGGAGTTTG-TA 2269
DB 661 TTCTCAAAACAAATCTCCACCAAGAACCCAGTCTCTGAATGGTGGAGTTTG-TA 720
QY 2270 GACAAACACCTTTG--CTGAAGAAATTCACACTCAGAAATCAGAAATCCCAAGGATGGAAAC 2327
DB 721 GACAAACACCTTTGCTGAAGAAATTCACACTCAGAAATCAGAAATCCCAAGGATGGAAAC 780
QY 2328 TCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGAAGAACAGATAAAGAGAAATCGGT 2387
DB 781 TCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGAAGAACAGATAAAGAGAAATCGGT 840
QY 2388 ATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCTTAAATTC 2447
DB 841 ATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCTTAAATTC 900
QY 2448 TGTATGTTTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATACAGAGGTATCCC 2507
DB 901 TGTATGTTTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATACAGAGGTATCCC 960
QY 2508 AGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTCTTAAATCAAA 2567
DB 961 AGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAGAAATTCCTTCTTAAATCAAA 1020
QY 2568 AACAAAAACAAAAACAAAAACAAATTCCTTAAATCAAAATTCCTTCTTAAATCAAA 2627
DB 1021 AACAAAAACAAAAACAAAAACAAATTCCTTAAATCAAAATTCCTTCTTAAATCAAA 1080
QY 2628 CTTCAATTTAGCAGTGATATGATCAATGAAGTGTGAAGCTTTGTAAGCTGGGGAATATT 2687
DB 1081 CTTCAATTTAGCAGTGATATGATCAATGAAGTGTGAAGCTTTGTAAGCTGGGGAATATT 1140
QY 2688 CCACCTGATATAGCCAGATTTCTACTGTATTCCTCAAAAGGCAATATTAAAGGTAGATAGA 2747
DB 1141 CCACCTGATATAGCCAGATTTCTACTGTATTCCTCAAAAGGCAATATTAAAGGTAGATAGA 1200
QY 2748 TGATTAGTAGTATTTGTTACACACTATTTGGAAATTTAGAGAACATACAGAGAAATTTA 2807
DB 1201 TGATTAGTAGTATTTGTTACACACTATTTGGAAATTTAGAGAACATACAGAGAAATTTA 1260
QY 2808 GGGGCTTAAACATTTAGCAGTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATATTTT 2867
DB 1261 GGGGCTTAAACATTTAGCAGTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATATTTT 1320
QY 2868 AAATGAATPACCAATTTAAATTTTATTTAGTATTTACCTGTTAAGAGATTTATTTAGTCTTTAA 2927
DB 1321 AAATGAATPACCAATTTAAATTTTATTTAGTATTTACCTGTTAAGAGATTTATTTAGTCTTTAA 1380
QY 2928 TTTTGTAGTTAAATTTCTCTGCTGATATATATAGAGAAATTTACTACTTTATGCTCTGC 2987
DB 1381 TTTTGTAGTTAAATTTCTCTGCTGATATATATAGAGAAATTTACTACTTTATGCTCTGC 1440
QY 2988 TCTCTAACTACATCTCAATCTCGAGCTCGCTGAGGTATATACACAGAGACATTTTGA 3047
DB 1441 TCTCTAACTACATCTCAATCTCGAGCTCGCTGAGGTATATACACAGAGACATTTTGA 1500
QY 3048 GGCAATTTGAAAAACCAACTACATCTCTCGGTGCTTTAGAGAGATCTGCTGCTCTCCCAAT 3107
DB 1501 GGCAATTTGAAAAACCAACTACATCTCTCGGTGCTTTAGAGAGATCTGCTGCTCTCCCAAT 1560
QY 3108 AAGCTTTTGTATCTGCCAGTGAATTTACTGTACTTCCCAATGATTCCTTTCTCGGTG 3167
DB 1561 AAGCTTTTGTATCTGCCAGTGAATTTACTGTACTTCCCAATGATTCCTTTCTCGGTG 1620

QY 3168 ATATCTGTGCTCTCATATAATTACTGAAGCTGCAATATTTTAGTAATACCTTCGGGATCA 3227
DB 1621 ATATCTGTGCTCTCATATAATTACTGAAGCTGCAATATTTTAGTAATA-CTTCGGGATCA 1679
QY 3228 CTGTCCCCCATCTTCCGTGTTTAGAGCAAAAGTGAAGAGTTTTAAAGGAGGAAGAAAGAA 3287
DB 1680 CTGTCCCCCATCTTCCGTGTTTAGAGCAAAAGTGAAGAGTTTTAAAGGAGGAAGAAAGAA 1739
QY 3288 CTGTCTTACACCACTTGAAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGATGTCCTC 3347
DB 1740 CTGTCTTACACCACTTGAAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGATGTCCTC 1799
QY 3348 TTTTGAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCAG 3407
DB 1800 TTTTGAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCAG 1859
QY 3408 TATTCTCAGGCTGAAATTTAAACCAACTATAGGCTTTTCTTTGGGATGATTTTCTAGTCT 3467
DB 1860 TATTCTCAGGCTGAAATTTAAACCAACTATAGGCTTTTCTTTGGGATGATTTTCTAGTCT 1919
QY 3468 TAAAGTTTGGGACATTTATAAACTTGAGTACATTTGTTGTACAGAGTTGATATTTCCAAAT 3527
DB 1920 TAAAGTTTGGGACATTTATAAACTTGAGTACATTTGTTGTACAGAGTTGATATTTCCAAAT 1979
QY 3528 TGTATGATGGAGGAGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGCTACTGCATTTATA 3587
DB 1980 TGTATGATGGAGGAGAGGTGCTTTAAGCTGTAGGCTTTTCTTTGCTACTGCATTTATA 2039
QY 3588 GAGATTTAGCTTTAATATTTTGTAGAGATGTAAGCACTTCTGCTTTCTTAGTCTTACCTA 3647
DB 2040 GAGATTTAGCTTTAATATTTTGTAGAGATGTAAGCACTTCTGCTTTCTTAGTCTTACCTA 2099
QY 3648 GTCTCAAAACATTTTATTTCAATAAAGATTTTAAATTTGAAAAAATAAAAAA 3705
DB 2100 GTCTCAAAACATTTTATTTCAATAAAGATTTTAAATTTGAAAAAATAAAAAA 2157

RESULT 12
ABI98973
ID ABI98973 standard; cDNA; 1754 BP.
XX
AC ABI98973;
XX
DT 21-FEB-2002 (first entry)
XX
DE Human cancer suppressor protein PP624 encoding cDNA.
XX
KW Human; cancer suppressor; disease; cancer; ss.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 444..1349
FT /*tag= a
FT /transl_except= (pos:104..106,aa:Xaa)
FT /transl_except= (pos:561..563,aa:Xaa)
FT /transl_except= (pos:570..572,aa:Xaa)
FT /product= "PP624"
FT /note= "Xaa = unknown"
XX
XX CN1313297-A.
XX
XX 19-SEP-2001.
XX
XX 09-MAR-2000; 2000CN-0111948.
XX
XX 09-MAR-2000; 2000CN-0111948.
XX
XX (SHAN-) SHANGHAI INST ONCOLOGY.
XX
XX Gu J, Yang S;
XX
XX WPI; 2002-042185/06.
DR

DR P-PSDB; ABB56420.
XX Human protein able to suppress growth of cancer cells and its coding
PT sequence -
XX
PS Claim 5; Page 21-22 Disclosure; 37pp; Chinese.
XX
CC The invention relates to novel human proteins (ABB56417-ABB56425) with
CC cancer suppressing function, the encoding polynucleotides
CC (AB19970-AB19978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
XX
SQ Sequence 1754 BP; 586 A; 377 C; 426 G; 361 T; 4 other;
Query Match 46.5%; Score 1721.4; DB 24; Length 1754;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1747; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 829 AGATGACCTGGAAATAGGCCCGAGTGTGTCATCTTCTACATTTGACTCGGAGAAAA 888
DB 1 AGATGACCTGGAAATAGGCCCGAGTGTGTCATCTTCTACATTTGACTCGGAGAAAA 60
QY 889 TGAGAGTAGACGAAATCTGAACTTCCACGCTCTCAGAACCTCTATAAAGGATCGAAT 948
DB 61 TGAGAGTAGACGAAATCTGAACTTCCACGCTCTCAGAACCTCTATAAAGGATCGAAT 120
QY 949 GGCCAACTACAGGCGAGCTGTGTCACAAACAAAGCAGCTCAACCAACTATACAAATGAGCT 1008
DB 121 GGCCAACTACAGGCGAGCTGTGTCACAAACAAAGCAGCTCAACCAACTATACAAATGAGCT 180
QY 1009 GAAAGCCAGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGAGAAATGTGCCCC 1068
DB 181 GAAAGCCAGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGAGAAATGTGCCCC 240
QY 1069 AGTCTGAGTGTGATCACCATCAGGAAGGGGAAAGATTTCTGCAATGAGAATAG 1128
DB 241 AGTCTGAGTGTGATCACCATCAGGAAGGGGAAAGATTTCTGCAATGAGAATAG 300
QY 1129 CTTGGCAGTCCGTTCCACCTCCGGAAGTACTCCGTTGACTCCAGGTTAAGAGTGA 1188
DB 301 CTTGGCAGTCCGTTCCACCTCCGGAAGTACTCCGTTGACTCCAGGTTAAGAGTGA 360
QY 1189 GGTTCACAGCCTGTCCATCCCAAGGCACCTAAGTCCAGAT - TCCAGAGCCTCCAGTCTTT 1247
DB 361 GGTTCACAGCCTGTCCATCCCAAGGCACCTAAGTCCAGATCTCCAGGCTCCAGTCTTT 420
QY 1248 CTGAAAGTTCTCTCCCAAGCAATGAAAGTTTCAGGCACCTGCAAGAGACCTGCG 1307
DB 421 CTGAAAGTTCTCTCCCAAGCAATGAAAGTTTCAGGCACCTGCAAGAGACCTGCG 480
QY 1308 TGGAAATGTCAGAGACAGTCTATCCATGAGGCTCTCTGGCCACCCAGCAGGTGTTTC 1367
DB 481 TGGAAATGTCAGAGACAGTCTATCCATGAGGCTCTCTGGCCACCCAGCAGGTGTTTC 540
QY 1368 ACATCAGCTGCTTCGTTGCTCTATTGCAACAACTCAGTCTAGGAACATATGCAAT 1427
DB 541 ACATCAGCTGCTTCGTTGCTCTATTGCAACAACTCAGTCTAGGAACATATGCAAT 600
QY 1428 CTTTACATGGAAGAACTATTGTAAGCCTCACTCAATCAACTCTTTAAATCTAAGGGCA 1487
DB 601 CTTTACATGGAAGAACTATTGTAAGCCTCACTCAATCAACTCTTTAAATCTAAGGGCA 660
QY 1488 ACTATGATGAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAATGAAA 1547
DB 661 ACTATGATGAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAATGAAA 720
QY 1548 ACGAAGAGATTTTGGAGACACAGCCAGCTTCAAAATCAAGGAGACCCCTCACAGCC 1607
DB 721 ACGAAGAGATTTTGGAGACACAGCCAGCTTCAAAATCAAGGAGACCCCTCACAGCC 780
QY 1608 CAGGGGTAGAAGATGCCCCCTATTGCTAAGGTGGGTGTCCTGCTGCAAGTATGAGAGCCA 1667

DB 781 CAGGGGTAGAAGATGCCCTTATGCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCA 840
QY 1668 AGGCCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGTGTGAGATCG 1727
DB 841 AGGCCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGTGTGAGATCG 900
QY 1728 CTTGSCCACCCTCCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGGATCAAAA 1787
DB 901 CTTGSCCACCCTCCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGGATCAAAA 960
QY 1788 TGTCAAAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCTCTGAGGATG 1847
DB 961 TGTCAAAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCTCTGAGGATG 1020
QY 1848 TCGATCTAGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCAT 1907
DB 1021 TCGATCTAGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCAT 1080
QY 1908 TCACCTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGTCCCCAC 1967
DB 1081 TCACCTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGTCCCCAC 1140
QY 1968 CTATCAGGAAGAGGCTGGAGCATGTTCAGAGCAGATGTGGAAGTCTGTGGGTGGAAGATTG 2027
DB 1141 CTATCAGGAAGAGGCTGGAGCATGTTCAGAGCAGATGTGGAAGTCTGTGGGTGGAAGATTG 1200
QY 2028 CAGAAAGGAAACAGTGGAAATGCAAGGCTTCTAACAAGATGGGAATGTGGGAAAAA 2087
DB 1201 CAGAAAGGAAACAGTGGAAATGCAAGGCTTCTAACAAGATGGGAATGTGGGAAAAA 1260
QY 2088 CAACCTGCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAGTAAAGAAAGTCTATA 2147
DB 1261 CAACCTGCAAAACAAAGAAATCTAAAGGAGAG - CAGGCAAGAGAGTAAAGAAAGTCTATA 1319
QY 2148 GTTTGAGATGAGAAATGAGAAATCTTTAGAAAAATGGTGCAGATCCGATGAAGATGATA 2207
DB 1320 GTTTGAGATGAGAAATGAGAAATCTTTAGAAAAATGGTGCAGATCCGATGAAGATGATA 1379
QY 2208 ACAGCTTCTCTCAACCAACAATCTCCACAGAACCAAGTCTCTGAAATGGTGCAGTTTG 2267
DB 1380 ACAGCTTCTCTCAACCAACAATCTCCACAGAACCAAGTCTCTGAAATGGTGCAGTTTG 1439
QY 2268 TAGACAACACCTTTCTCTGAAAGAAATCTACTCTCAGAAATCCCAAGATGTGGAAAC 2327
DB 1440 TAGACAACACCTTTCTCTGAAAGAAATCTACTCTCAGAAATCCCAAGATGTGGAAAC 1499
QY 2328 TCTGGAGGGAGAAAGTGTCAAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATCGGT 2387
DB 1500 TCTGGAGGGAGAAAGTGTCAAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATCGGT 1559
QY 2388 ATTATGATGAGATGAGATGAAGAGTCAAAAATTCGAAATGATGCTGGGCTTAAATTTCA 2447
DB 1560 ATTATGATGAGATGAGATGAAGAGTCAAAAATTCGAAATGATGCTGGGCTTAAATTTCA 1619
QY 2448 TCTTAGTGTAGCAGGACCTGCCCTTTGTCAAAATGTATGTCACATAGCAGGATATCCC 2507
DB 1620 TCTTAGTGTAGCAGGACCTGCCCTTTGTCAAAATGTATGTCACATAGCAGGATATCCC 1679
QY 2508 AGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAGAAAGTCTCTTTAAAAATCAAA 2567
DB 1680 AGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAGAAAGTCTCTTTAAAAATCAAA 1739
QY 2568 AACAAAAA 2582
DB 1740 AAAAAAAAAA 1754

RESULT 13
AA159955
ID AA159955 standard; cDNA; 1713 BP.
XX
AC AA159955;

XX
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 3944.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.
XX
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX DR WPI: 2001-442253/47.
XX DR P-PSDB; AAM40799.
XX
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX PS Claim 1; SEQ ID NO 3944; 100789p; English.
XX
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX CC the encoded polypeptides (AAM38642-AAM4213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX
XX SQ Sequence 1713 BP; 549 A; 401 C; 417 G; 346 T; 0 other;

Query Match 43.0%; Score 1595; DB 22; Length 1713;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 1642; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

Qy 22 TAGCAGCTTGTCGCGACAGGTGCGCTAGGTAGAGCGCGGACCTGTGACAGGGCTGGT 81
Db 6 TAGCAGCTTGTCGCGACAGGTGCGCTAGGTAGAGCGCGGACCTGTGACAGGGCTGGT 65

Qy 82 AGCAGCGCAGAGGAAAGCGCGCTTTTAGCCAGGATTTTCAGTGTGTAGACAAGATGGA 141
Db 66 AGCAGCGCAGAGGAAAGCGCGCTTTTAGCCAGGATTTTCAGTGTGTAGACAAGATGGA 125

Qy 142 ATCATCTCCATTTAATAGAGCGCAATGGACCTCACTATCATTCATTGAGGGTAACAGCCAAAGA 201
Db 126 ATCATCTCCATTTAATAGAGCGCAATGGACCTCACTATCATTCATTGAGGGTAACAGCCAAAGA 185

Qy 202 ACTTCTCTTGTCAACAAGAACAAAGTCAATCGGCTATTGTGGAATATTCTTCAAGTACCA 261
Db 186 ACTTCTCTTGTCAACAAGAACAAAGTCAATCGGCTATTGTGGAATATTCTTCAAGTACCA 245

Qy 262 GAAAGCAGCTGAAGAAACAAACATGAGAGAGAGAGAGTAAACACGAAATCTCTCCCA 321
Db 246 GAAAGCAGCTGAAGAAACAAACATGAGAGAGAGAGAGTAAACACGAAATCTCTCCCA 305

Qy 322 GCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGGAAACCCAGGGCTGGG 381
Db 306 GCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGGAAACCCAGGGCTGGG 365

Qy 382 AGCAGAGTCTCACACAGACTCTCTACGGAAACAGCAGCACTGAGATTAGGCACAGAGCAGA 441
Db 366 AGCAGAGTCTCACACAGACTCTCTACGGAAACAGCAGCACTGAGATTAGGCACAGAGCAGA 425

Qy 442 CCATCTCTCTGCTGAAGTGCACAAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAAGA 501
Db 426 CCATCTCTCTGCTGAAGTGCACAAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAAGA 485

Qy 502 ACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAGCCCTCGTTTCAGGGTGCATA 561
Db 486 ACAAATCCACCCAGATCTAGACTCAGGGCACTCTCTGAAGCCCTCGTTTCAGGGTGCATA 545

Qy 562 TCCCCACATCAAGGACGGTGAAGATCTTAAAGACCACTCAACAGAAAGTAAAAAAATGGA 621
Db 546 TCCCCACATCAAGGACGGTGAAGATCTTAAAGACCACTCAACAGAAAGTAAAAAAATGGA 605

Qy 622 AATTTGCTAGAGAAATCCAGGCATGAAGTGAAGAAATCAGAAATCAGTGAAGAAACACAGA 681
Db 606 AATTTGCTAGAGAAATCCAGGCATGAAGTGAAGAAATCAGAGATCAGTGAAGAAACACAGA 665

Qy 682 TGGTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAAGATGTTTGA 741
Db 666 TGGTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAAGATGTTTGA 725

Qy 742 GAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTGCAGGTGGAAG 801
Db 726 GAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTGCAGGTGGAAG 785

Qy 802 GAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCACAGGTGAGTTGTC 861
Db 786 GAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCACAGGTGAGTTGTC 845

Qy 862 ATCTTCTACATTTGACTCGGAGAAATAGAGTGAAGTGAAGTCTGGAACCTTCCACGCT 921
Db 846 ATCTTCTACATTTGACTCGGAGAAATAGAGTGAAGTGAAGTCTGGAACCTTCCACGCT 905

Qy 922 CTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAGGACGCTGTGTCACAAACAAG 981
Db 906 CTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAGGACGCTGTGTCACAAACAAG 965

Qy 982 CAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAA 1041
Db 966 CAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAA 1025

Qy 1042 AATGAGCAAAAGGAGATGTGCCCCCAGGCTCTGAGGTCTGATCAACCCATCAGGAAGG 1101
Db 1026 AATGAGCAAAAGGAGATGTGCCCCCAGGCTCTGAGGTCTGATCAACCCATCAGGAAGG 1085

Qy 1102 GAAAAGATTTCTGCAAAATGAGATAGCTCGCAGTCCGTTCCACCCCTGCCGAAGATGA 1161
Db 1086 GAAAAGATTTCTGCAAAATGAGATAGCTCGCAGTCCGTTCCACCCCTGCCGAAGATGA 1145

Qy 1162 CTCCTGACTCCCAGGTTAAGAGTGAAGTTCACAGGCTGTCCATCCCAAGCCACTAAG 1221
Db 1146 CTCCTGACTCCCAGGTTAAGAGTGAAGTTCACAGGCTGTCCATCCCAAGCCACTAAG 1205

QY 1222 TCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCCTCCCAAGCAATGAAGAGTT 1281
DB 1206 TCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCCTCCCAAGCAATGAAGAGTT 1265
QY 1282 TCAGGCACCTGCAAGAGACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAAATGGAGCG 1341
DB 1266 TCAGGCACCTGCAAGAGACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAAATGGAGCG 1325
QY 1342 TCTCTTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTCCGTTGCTCTTATTCGAACAA 1401
DB 1326 TCTCTTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTCCGTTGCTCTTATTCGAACAA 1385
QY 1402 CAAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTT 1461
DB 1386 CAAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTT 1445
QY 1462 CAATCAACTCTTTAAATCTAAGGSCAACTATGATGAAGGCTTTGGGCACAGACCAACAA 1521
DB 1446 CAATCAACTCTTTAAATCTAAGGSCAACTATGATGAAGGCTTTGGGCACAGACCAACAA 1505
QY 1522 GGATCTATGGGCAAGCAAAATGMAACGAAGAGATTTTGGAGAGACCCAGCCAGTTGC 1581
DB 1506 GGATCTATGGGCAAGCAAAATGMAACGAAGAGATTTTGGAGAGACCCAGCCAGTTGA 1565
QY 1582 AAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGA-TGCCCTTATTGCTAAGGTGG 1640
DB 1566 AAATCAAGGAGAGACCCCTCAAAAGCCAGGGGTGAAGATTGCCCGAGTTGCTAAGGTGG 1625
QY 1641 GTGTCTCTGGTGAAGATATGGAAGCCAAAGCCCTCCTCTCAGCAGGAGAGGAAGACAAGC 1700
DB 1626 TTGTCTCTGGTGAAGATATGGAAGCCAAAGCCCTCCTCTCAGCAGGAGAGGAAGACAAGC 1685
QY 1701 CAGCTGAAACCAAGAGCTGAGG 1723
DB 1686 CAGCTGGAACCCGAACTTAGGG 1708

RESULT 14
AAH18125
ID AAH18125 standard; cDNA; 2207 BP.
XX
AC AAH18125;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17991.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
XX
XX EP1074617-A2.
PD
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -
XX Claim 8; SEQ ID 17991; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2207 BP; 692 A; 438 C; 512 G; 565 T; 0 Other;

Query Match 42.3%; Score 1568.4; DB 22; Length 2207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1001 AATGAGCTGAAAGCCAGTGGTGGCCAAATCAAAATTCATAAAATGGACAAAGAGAAAT 1060
DB 638 AATGAGCTGAAAGCCAGTGGTGGCCAAATCAAAATTCATAAAATGGACAAAGAGAAAT 697
QY 1061 GTGCCCCCAGGTCTCTGAGTCTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAAT 1120
DT 698 GTGCCCCCAGGTCTCTGAGTCTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAAT 757
QY 1121 GAGATAGCCTGGAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTGACTTCCAGGTT 1180
DB 758 GAGATAGCCTGGAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTGACTTCCAGGTT 817
QY 1181 AAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
DB 818 AAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 877
QY 1241 AGTCTTTCTGAAAGTTCCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 1300
DB 878 AGTCTTTCTGAAAGTTCCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 937
QY 1301 ACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAATGGAGCGTCTCTTGGCCCAACAGCAG 1360
DB 938 ACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAATGGAGCGTCTCTTGGCCCAACAGCAG 997
QY 1361 GTGTTTACATCAGCTGCTTCCGTTGCTCTTATTCGAACCAACAACTCAGTCTAGGAACA 1420
DB 998 GTGTTTACATCAGCTGCTTCCGTTGCTCTTATTCGAACCAACAACTCAGTCTAGGAACA 1057
QY 1421 TATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTTTTAAATCT 1480
DB 1058 TATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTTTTAAATCT 1117
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAGAGGATCTATGGCAGCAAA 1540
DB 1118 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAGAGGATCTATGGCAGCAAA 1177
QY 1541 AATGAAGCAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGAGACCCCT 1600
DB 1178 AATGAAGCAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGAGACCCCT 1237

QY 1601 CACAGCCAGGGGTAGAGATGCCCTTATTGCTAGGTGGGTGCTCTGGCTGCAAGTATG 1660
D 1238 CACAGCCAGGGGTAGAGATGCCCTTATTGCTAGGTGGGTGCTCTGGCTGCAAGTATG 1297
QY 1661 GAAGCCAGGGCTCTCTCAGCAGAGAGAGAGACAGCAGCTGAAACCAAGAGCTG 1720
D 1298 GAAGCCAGGGCTCTCTCAGCAGAGAGAGAGAGACAGCAGCTGAAACCAAGAGCTG 1357
QY 1721 AGGATCGCTGGCCACCCCTCACTGAACTTGGAGTTTCAGGAAGTCTGGAGCAAGGG 1780
D 1358 AGGATCGCTGGCCACCCCTCACTGAACTTGGAGTTTCAGGAAGTCTGGAGCAAGGG 1417
QY 1781 ATCAAAATGTCAAGCCCAAAATGCGCTCTCAAGCAGAAATCAGCAAGCCCGAAGTTCT 1840
D 1418 ATCAAAATGTCAAGCCCAAAATGCGCTCTCAAGCAGAAATCAGCAAGCCCGAAGTTCT 1477
QY 1841 GAGGATGTCGATCTAGATCTGAAGAGCTAAGAGATCTTCTCACTGAAGGAAGAGC 1900
D 1478 GAGGATGTCGATCTAGATCTGAAGAGCTAAGAGATCTTCTCACTGAAGGAAGAGC 1537
QY 1901 CGCCATTCACTAGCAGCTTCACTTCAAGCAGCTCTGTCAAGAGCCCAAAACTGTG 1960
D 1538 CGCCATTCACTAGCAGCTTCACTTCAAGCAGCTCTGTCAAGAGCCCAAAACTGTG 1597
QY 1961 TCCCACTATCAGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGATCTGTGGGTGA 2020
D 1598 TCCCACTATCAGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGATCTGTGGGTGA 1657
QY 2021 AGATTGTCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTG 2080
D 1658 AGATTGTCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAGAGAAATGGGAATGTG 1717
QY 2081 GGAAAAACAACCTGGCAAAACAAGATCTTAAAGAGAGACAGGGAAGAGAAAGTGAAGAA 2140
D 1718 GGAAAAACAACCTGGCAAAACAAGATCTTAAAGAGAGACAGGGAAGAGAAAGTGAAGAA 1777
QY 2141 GGTATAGTTGGAGATGAGAAATGAGAAATCTTGTAGAAATGGTGCAGACTCCGATGAA 2200
D 1778 GGTATAGTTGGAGATGAGAAATGAGAAATCTTGTAGAAATGGTGCAGACTCCGATGAA 1837
QY 2201 GATGATACAGCTTCTCAACACAACTCTCCCAAGAACCCCAAGTCTCTGAATTTGGTGC 2260
D 1838 GATGATACAGCTTCTCAACACAACTCTCCCAAGAACCCCAAGTCTCTGAATTTGGTGC 1897
QY 2261 AGTTTGTAGACAACCTTTGCTGAAGATTTCACTACTCAGATCAGAAATCCAGGAT 2320
D 1898 AGTTTGTAGACAACCTTTGCTGAAGATTTCACTACTCAGAAATCAGAAATCCAGGAT 1957
QY 2321 GTGAAACTCTGGAGGGAGAGAGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2380
D 1958 GTGAAACTCTGGAGGGAGAGAGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2017
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATTTGCAATGATGCTGGGCTT 2440
D 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATTTGCAATGATGCTGGGCTT 2077
QY 2441 AAATTCATGTTAGTGTAGCGAGCCCTGCGCTTTGTCAAAATGTGATGCACATAAGCAG 2500
D 2078 AAATTCATGTTAGTGTAGCGAGCCCTGCGCTTTGTCAAAATGTGATGCACATAAGCAG 2137
QY 2501 GTATCCAGCATGAATATGTAATTTACTTGGAGTAACCTTTGGAAAGAAATTCCTTCTTAA 2560
D 2138 GTATCCAGCATGAATATGTAATTTACTTGGAGTAACCTTTGGAAAGAAATTCCTTCTTAA 2197
QY 2561 AATCAAAAC 2570
D 2198 AATCAAAAC 2207

RESULT 15
AAS18588
ID AAS18588 standard; cDNA; 1567 BP.
XX

AC AAS18588;
XX 12-MAR-2002 (first entry)
XX cDNA encoding human sterol regulatory element binding protein 3.
DE Human; sterol regulatory element binding protein 3; hSREBP-3;
XX hypothalamus; ss.
KW Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 169..154
FT /tag= a
FT /product= "Sterol regulatory element binding protein 3,
FT hSREBP-3"
XX
XX CN1309182-A.
XX 22-AUG-2001.
XX 17-FEB-2000; 2000CN-0111698.
XX 17-FEB-2000; 2000CN-0111698.
XX (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.
XX Li Y, Xu S, Ren S;
XX WPI; 2002-011822/02.
DR P-PSDB; AAU10979.
XX
XX Cholesterol regulatory factor binding protein and its coding sequence .
XX Claim 1; Page 22; 27pp; Chinese.
XX
CC The invention relates to a novel human sterol regulatory element binding
CC protein 3 (hSREBP-3) expressed in human normal hypothalamic tissue and
CC its coding sequence. Also described is the process for preparing the
CC protein and nucleic acid sequence, and the method for detecting hSREBP-3
CC nucleic acid sequence and polypeptides. The present sequence represents
CC the coding sequence of human hSREBP-3 as described in the invention.
XX
SQ Sequence 1567 BP; 513 A; 326 C; 391 G; 337 T; 0 other;
Query Match 38.7%; Score 1433; DB 24; Length 1567;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGGAGAT 1060
D 127 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGGAGAT 186
QY 1061 GTGCCCCAGGTCCTGAGTCTGCATCACCCTCAGGAAGGGGAAAGATTTCTGCAAT 1120
D 187 GTGCCCCAGGTCCTGAGTCTGCATCACCCTCAGGAAGGGGAAAGATTTCTGCAAT 246
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGCAAGATGACTCCCGTACTCCAGGTT 1180
D 247 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGCAAGATGACTCCCGTACTCCAGGTT 306
QY 1181 AAGAGTGAAGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTTCCAGAGCCTCC 1240
D 307 AAGAGTGAAGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTTCCAGAGCCTCC 366
QY 1241 AGTCTTCTGAAAGTTCTCTCCCAAGCAAGCAATGAAGAGTTTCAGGCACCTGCAGAGAG 1300
D 367 AGTCTTCTGAAAGTTCTCTCCCAAGCAAGCAATGAAGAGTTTCAGGCACCTGCAGAGAG 426
QY 1301 ACCTCGTGGAAATGTGCAAGACAGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAG 1360
D 427 ACCTCGTGGAAATGTGCAAGACAGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAG 486

QY 1361 GTGTTTCACATCAGCTGCTTCCGTTGCTCCTATTGCAACAACAACTCAGTCTAGGAACA 1420
Db GTGTTTCACATCAGCTGCTTCCGTTGCTCCTATTGCAACAACAACTCAGTCTAGGAACA 546
QY 1421 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCAGTTCATCAACTCTTTAAATCT 1480
Db TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCAGTTCATCAACTCTTTAAATCT 606
QY 1481 AAGGCAACTATGATGAAGCTTTGGGCAACAGACCACCAAGGATCTATGGGCAAGCAAA 1540
Db AAGGCAACTATGATGAAGCTTTGGGCAACAGACCACCAAGGATCTATGGGCAAGCAAA 666
QY 1541 AATGAAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGGAGACCCCT 1600
Db AATGAAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGGAGACCCCT 726
QY 1601 CACAGCCCAAGGGTAGAAGATGCCCTATTGCTAAGGTGGGTGTCTCTGGCTGCAAGTATG 1660
Db CACAGCCCAAGGGTAGAAGATGCCCTATTGCTAAGGTGGGTGTCTCTGGCTGCAAGTATG 786
QY 1661 GAAGCCAAAGGCTCTCTCAGCAGGAGAAAGAGACAAGCCAGCTGAAACCAAGAAAGCTG 1720
Db GAAGCCAAAGGCTCTCTCAGCAGGAGAAAGAGACAAGCCAGCTGAAACCAAGAAAGCTG 846
QY 1721 AGGATCGCTGGCCACCCCTGAACTTGAAGTTTCAGAAAGTGCCTTGGAGGAAGGG 1780
Db AGGATCGCTGGCCACCCCTGAACTTGAAGTTTCAGAAAGTGCCTTGGAGGAAGGG 906
QY 1781 ATCAAAAATGTCAAAGCCCAATGGCCTCTGAAAGACGAAATCAGAAAGCCCGAAGTTCT 1840
Db ATCAAAAATGTCAAAGCCCAATGGCCTCTGAAAGACGAAATCAGAAAGCCCGAAGTTCT 966
QY 1841 GAGATGTGATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1900
Db GAGATGTGATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1026
QY 1901 CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTGTCAGAGCCCAAAAACCTGTG 1960
Db CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTGTCAGAGCCCAAAAACCTGTG 1086
QY 1961 TCCCCACCTATCAGGAAGGCTGGAGCATGTGAGCAGAGTCAAGAGTCTGTGGGTGGA 2020
Db TCCCCACCTATCAGGAAGGCTGGAGCATGTGAGCAGAGTCAAGAGTCTGTGGGTGGA 1146
QY 2021 AGAGTTGCAAAAAGGAACAAAGTGGAAAATGCCAAGGCTTCTAAGGAAGATGGGAATGTG 2080
Db AGAGTTGCAAAAAGGAACAAAGTGGAAAATGCCAAGGCTTCTAAGGAAGATGGGAATGTG 1206
QY 2081 GGAATAACACCTGGGCAACAAAGATCTAAGGAGAGACAGGGAAGAGAGTAAGGAA 2140
Db GGAATAACACCTGGGCAACAAAGATCTAAGGAGAGACAGGGAAGAGAGTAAGGAA 1266
QY 2141 GGTCTAGTTTGGAGATGGAGATGAGATCTTTGTAGAAAATGGTGCAGACTCCGATGAA 2200
Db GGTCTAGTTTGGAGATGGAGATGAGATCTTTGTAGAAAATGGTGCAGACTCCGATGAA 1326
QY 2201 GATGATAACAGCTTCTCAAAACAAATCTCCACAAGAACCCCAAGTCTCTGAATGGTCG 2260
Db GATGATAACAGCTTCTCAAAACAAATCTCCACAAGAACCCCAAGTCTCTGAATGGTCG 1386
QY 2261 AGTTTGTAGACAACACCTTTGCTGAAGAAATTCACCTACTCAGAAATCCAGGAT 2320
Db AGTTTGTAGACAACACCTTTGCTGAAGAAATTCACCTACTCAGAAATCCAGGAT 1446
QY 2321 GTGGAACTCTGGGAGGAGAGTGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 2380
Db GTGGAACTCTGGGAGGAGAGTGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 1506
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAGAGTGAATAATTGCAATGATGCTGGGCCCTT 2440
Db AATCGGTATTATGATGAGGATGAGGATGAGAGTGAATAATTGCAATGATGCTGGGCCCTT 1566
QY 2441 A 2441

Db 1567 A 1567

Search completed: January 6, 2004, 10:12:20
Job time : 910 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:52:31 ; Search time 7136 Seconds
(without alignments)

12531.044 Million cell updates/sec

Title: US-09-890-549-16

Perfect score: 3705

Sequence: 1 ggcgcaggagcagtaggtg.....atttgaaaaaaaaaaaaaa 3705

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba:**

2: em_esthm:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_mam:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1557.4	42.0	3242	11 AK049350	AK049350 Mus muscu
2	1403.2	37.9	3108	11 AK031698	AK031698 Mus muscu
3	1401.6	37.8	3100	11 AK028186	AK028186 Mus muscu
4	1400	37.8	3160	11 AK085065	AK085065 Mus muscu

C	5	949.6	25.6	1014	9	AL578990
	6	927.6	25.0	1045	9	AL556858
	7	898.2	24.2	1075	12	BM909158
	8	806.8	21.8	880	12	BI871114
	9	802.4	21.7	884	14	CD171644
	10	787.8	21.3	1142	12	BM914155
	11	775.4	20.9	810	12	BM721370
	12	774	20.9	785	10	BG740815
C	13	766.4	20.7	777	28	AQ114676
	14	766	20.7	828	12	BI457843
	15	764.2	20.6	781	10	BG676749
	16	744	20.1	759	12	BM977831
	17	732.4	19.8	759	14	CA447584
	18	731.8	19.8	767	12	BM981550
C	19	731.4	19.7	745	14	CB055210
	20	728	19.6	768	10	BG476496
	21	726.8	19.6	989	12	BM552304
	22	724.8	19.6	728	13	B0631767
	23	722.8	19.5	839	10	BG697786
	24	721	19.5	971	10	BE410108
C	25	719.4	19.4	725	13	BU740444
	26	717.8	19.4	724	12	BM980979
	27	713.2	19.2	742	12	BI766182
	28	711.2	19.2	855	12	BG777288
	29	707.8	19.1	787	13	BU861039
C	30	707	19.1	711	14	CA447848
	31	706	19.1	931	12	BM449793
	32	703.2	19.0	918	13	BU850203
	33	698.4	18.9	719	14	CA412726
	34	697.8	18.8	745	10	BG698125
	35	697.2	18.8	724	14	CB055211
	36	696	18.8	847	13	BU603212
C	37	694	18.7	843	9	AA909018
	38	692.6	18.7	951	10	BF528886
	39	692	18.7	715	9	AL048161
	40	687.6	18.6	770	12	BI089829
C	41	685.4	18.5	695	14	CA420719
	42	685.2	18.5	701	14	CD365699
C	43	683.4	18.4	752	10	BE897604
	44	683	18.4	767	14	CB988778
	45	682.8	18.4	1120	10	BG120125

ALIGNMENTS

RESULT 1
AK049350
LOCUS
DEFINITION
Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone: C330026L07 product: epithelial protein lost in neoplasm, full
insert sequence.
ACCESSION
AK049350
VERSION
AK049350.1 GI:26340071
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci, P. and Hayashizaki, Y.
AUTHORS
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

AK049350 3242 bp mRNA linear HTC 05-DEC-2002

<p>REFERENCE</p> <p>JOURNAL</p> <p>MEDLINE</p> <p>PUBMED</p> <p>REFERENCE</p> <p>AUTHORS</p>	<p>Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohgami,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsushita,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.</p> <p>RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer</p> <p>Genome Res. 10 (11), 1757-1771 (2000)</p> <p>20530913</p> <p>11076861</p> <p>4</p>	<p>Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casasvanti,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikolaev,I., Peole,G., Komiya,H., Quackenbush,J., Schriml,L.M., Stauber,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sekai,K., Okido,T., Furuno,M., Anon.H., Baldarelli,R., Barsch,G., Blake,J., Boffelli,D., Bojunga,N., Carlini,P., de Bonaldo,M.F., Brownstein,M.J., Bult.C., Fletcher,C., Fujita M., Gariboldi,M., Gustincich,S., Hill.D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Wittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsuki,S. and Hayashizaki,Y.</p> <p>Functional annotation of a full-length mouse cDNA collection</p> <p>Nature 409 (6921), 685-690 (2001)</p> <p>21085660</p> <p>11217851</p> <p>5</p>	<p>The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II team.</p> <p>Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs</p> <p>Nature 420, 563-573 (2002)</p> <p>6 (bases 1 to 3242)</p> <p>Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hitaoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kusuda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sugabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.</p> <p>Direct Submission</p> <p>Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site for further details.</p> <p>URL:http://genome.gsc.riken.go.jp/</p> <p>URL:http://fantom.gsc.riken.go.jp/</p> <p>Location/Qualifiers</p>	<p>TITLE</p> <p>JOURNAL</p> <p>MEDLINE</p> <p>PUBMED</p> <p>REFERENCE</p> <p>AUTHORS</p>	<p>Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohgami,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsushita,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.</p> <p>RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer</p> <p>Genome Res. 10 (11), 1757-1771 (2000)</p> <p>20530913</p> <p>11076861</p> <p>4</p>	<p>Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casasvanti,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikolaev,I., Peole,G., Komiya,H., Quackenbush,J., Schriml,L.M., Stauber,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sekai,K., Okido,T., Furuno,M., Anon.H., Baldarelli,R., Barsch,G., Blake,J., Boffelli,D., Bojunga,N., Carlini,P., de Bonaldo,M.F., Brownstein,M.J., Bult.C., Fletcher,C., Fujita M., Gariboldi,M., Gustincich,S., Hill.D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Wittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsuki,S. and Hayashizaki,Y.</p> <p>Functional annotation of a full-length mouse cDNA collection</p> <p>Nature 409 (6921), 685-690 (2001)</p> <p>21085660</p> <p>11217851</p> <p>5</p>	<p>The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II team.</p> <p>Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs</p> <p>Nature 420, 563-573 (2002)</p> <p>6 (bases 1 to 3242)</p> <p>Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hitaoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kusuda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sugabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.</p> <p>Direct Submission</p> <p>Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site for further details.</p> <p>URL:http://genome.gsc.riken.go.jp/</p> <p>URL:http://fantom.gsc.riken.go.jp/</p> <p>Location/Qualifiers</p>	<p>TITLE</p> <p>JOURNAL</p> <p>MEDLINE</p> <p>PUBMED</p> <p>REFERENCE</p> <p>AUTHORS</p>	<p>Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohgami,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsushita,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.</p> <p>RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer</p> <p>Genome Res. 10 (11), 1757-1771 (2000)</p> <p>20530913</p> <p>11076861</p> <p>4</p>	<p>Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casasvanti,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikolaev,I., Peole,G., Komiya,H., Quackenbush,J., Schriml,L.M., Stauber,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sekai,K., Okido,T., Furuno,M., Anon.H., Baldarelli,R., Barsch,G., Blake,J., Boffelli,D., Bojunga,N., Carlini,P., de Bonaldo,M.F., Brownstein,M.J., Bult.C., Fletcher,C., Fujita M., Gariboldi,M., Gustincich,S., Hill.D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Wittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsuki,S. and Hayashizaki,Y.</p> <p>Functional annotation of a full-length mouse cDNA collection</p> <p>Nature 409 (6921), 685-690 (2001)</p> <p>21085660</p> <p>11217851</p> <p>5</p>	<p>The FANTOM Consortium and the R</p>
--	--	--	--	--	--	--	--	--	--	--	--


```
Db      436  ACGGAGAGCCAAATAATGAAATCTGTCTGGAGATTCCAGGCATGAAGCAGAGAAGCCA 495
QY      662  GAAATCAGTGAAGAACACACAGATGCTTTGGGCAAAATAGAGAAATATAATGTTCCCGTGAAC 721
Db      496  GAGACGAGCGAAACACACAGAACTTCAGGCAAAATAGAGAAATACAAGTTTCCACTGAAT 555
QY      722  AGCTTTAGATGATGTTGTAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAA 781
Db      556  AGACTCAAGATGATGTTGTAGAAAGGTGAACACAAACCAACCAAGAGTCTCTGACCCCAA 615
QY      782  AGCCCAAGTGAAGTGAAGGAGAGATCTCTGAAACAGCTTATTTCTAGATGAAGCTGGAA 841
Db      616  AGCCGAAATGGGGTGAAGGAGGCTCTCTGAAACCAACTGTTCCCTGGATGACTGGAA 675
QY      842  ATAGCCCGAGGTGAGTTGTCATCTTCTACATTTGACTCGGAGAAATAGAGATAGACGA 901
Db      676  ATAGTGGGGTCAATTTGCTCTCTCTGCTCAATCACTCGGAGAAACAGAGATGAAGG 735
QY      902  AATCTGGAACCTCCACGGCTCTCAGAAACCTCTATAAGAGATCGAATGGCCAAAGTACCAG 961
Db      736  AATCTGAGCTGCCACGGCTCTCAGAAACCTCTCATAAGAGCCGATGGCCAAAGTACCAG 795
QY      962  GCAGCTGTGTCAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 1021
Db      796  GCTGCAGTGTCCAAAGCAGAGAGCCGACAGCTACACAAATGAGCTGAAACCCAGT--- 852
QY      1022  GCGCAATCAAAATTCATAAAATGAGCAAAAGCAGATGTGCCCCAGGTCTCTGAGGTC 1081
Db      853  ---GAAAGCAAAATCTATAATGGAACAGAGAGAGATGTGCCCCAGGTCTCCGAGGCC 909
QY      1082  TGCATCACCCATCAGGAAGGGGAAAGATTCTTCAAAATGAGATAGCCTGGCAGTCCGT 1141
Db      910  TGCAGCTCCATCAGGAAGGAAGAGGTTTCTACACTGAGAAATAGTCTAGTGGGCCCTC 969
QY      1142  TCCACCCCTCGCGAAGATGACTCCCGTGAATCCAGGTTAAGAGTGAAGTTCAACAGCCT 1201
Db      970  TCTCTCCCTGCTGAAGATGACACTGTAACTCCACAGTGAAGAGCGGCCAGCAGCCT 1029
QY      1202  GTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTTCCAGTCTTCTGAAAGTCTCCT 1261
Db      1030  ATGCACCCCTAAGCGCTGAGTCTCAGCCAGGACCTTCCAGTCTTCCGGAAGTCTCCT 1089
QY      1262  CCCAAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGCAAG 1321
Db      1090  TCCAAACACGGAAGAGTTTCAGCGCCGCGCAAGAGAGAGCTGCGTGGAGTGTCAAG 1149
QY      1322  ACAGTCTATCCAAATGGAGCGTCTCTTGGCCAAACAGCAGAGTGTTCACATCAGCTCCTC 1381
Db      1150  ACGGTGTACCCCATGGAACGGCTCCTGGCCAAACAGCAGAGTGTTCACATCAGCTGTTTC 1209
QY      1382  CGTGTCTCTATTGCAACAAACAACTCAGTCTAGGAACATATGATCTTTTACATGGAAGA 1441
Db      1210  CGATGCTCTCTACTGCAACAAACAAAGCTCAGTCTAGGAACATATGCACTCTTACATGGA 1269
QY      1442  ATCTATTGTAAGCCTCAGTCTCACTCACTCTTTAAATCTAAGGGCAACTATGATGAAGC 1501
Db      1270  ATCTCTGTAAGCCTCAGTCTCACTCACTCTTTAAATCTAAGTCCAAAGGCAACTATCAGAG 1329
QY      1502  TTTGGGCACAGACACACAGGATCTATGGGCAAGCAAAATGAAACGAAAGAGATTG 1561
Db      1330  TTTGGGCACAGCAGATGAAGATCTGTGGCAAGCAAGAGGCAATGAGGAGACTTTG 1389
QY      1562  GAGAGACAGCCAGCTTGCATTAAGAGGAGAGCCCTCACAGCCGAGGGGTAGAAGAT 1621
Db      1390  GGGAGACAGCCAGCCGCTTAATGAGGGGAGAGCCCTCATAGCCGGGTGTAGAAGAT 1449
QY      1622  GCCCTATTGCTAAGTGGGTGCTCGGTGCAAGTATGAAGCAAGCCCTCTCTCTCAG 1681
Db      1450  GCCCCCATCGCAAGGTGCGCGTGTGGGCGAAGTATGGAAGCAAGCCCTCTCTCTCAG 1509
QY      1682  CAGGAGAAGGAACAAGCCAGCTGAACCAAGAGAGCTGAGATCGCTGCGCACCCCTCC 1741
```

```
Db      1510  AGGAGAGGGAGGATAGCCTCGGAGACCAAGAGCTGAGGATCGCCTGGCCTCCGCCA 1569
QY      1742  ACTGAATCTTGAAGATTTCAGGAAGTGCCTTCAGGAAGAGGATCAAAATGTCAAAGCCCAAA 1801
Db      1570  GCCAGCTGGCGGTTCCGGAAGTGCCTTCAGGAAGAGGATCAAGATCAAGAGCCCAAG 1629
QY      1802  TGGCTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCCTGAGGATGTCGATCTAGATCTG 1861
Db      1630  TGGCTCTCGGAGGATGAGTCTGCAAGACGAGAGGCCCGGAGGATGTAGATCTGGACCTG 1689
QY      1862  AAGAAGCTAAGACGATCTTCTTCACTCAAGGAAGAGCGGCCCATTCCTCTAGCAGCT 1921
Db      1690  AAGAAGCTGCGGGGCTTTCGTCCCTCAAGGAGAGGAGCGGCCGTTTACGGTAGCAGCT 1749
QY      1922  TCATTTCAAAGCACCTCTGTCAAGAGCCCAAAATCTGTGTCCCCACCTATCAGGAAGGC 1981
Db      1750  TCCTTTCAAGACCTCTTCCATCAGAGCCCAAGCCCTCTTTCGCCATCTCTCAGGAAGGT 1809
QY      1982  TGAAGCATGTCAAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGATTGCA---GAAAGGAAA 2038
Db      1810  TGAAGGAGTCCGAGCAGAGTGAAGAGTTCGAGGAGGATAGCGACGATGGAAGGAAA 1869
QY      2039  CAAGTGGAAATGCGCAAGGCTTCTAAGAAAGATGGAAATGTGGGAAAAACAACCTCGCAA 2098
Db      1870  CAACCGAAATGCGCAAGGCTCTGCGGAGAAAGAGATGTGGGAAAAATCAAGCTGGCAG 1929
QY      2099  ACAAAGAAATCAAGAGAGAGACAGGAAAGAGAGTAAAGGAAGGTCAAGTTTGGAGATG 2158
Db      1930  -----GGCAGGAAGTCCGAGGAGCAAGACCGTAGTTCTTTTGGAGT 1974
QY      2159  GAGAAATGAGAAATCTTTAGAAAATGGTGAGACTCCGATGAAGATGATAACAGCTTCCTC 2218
Db      1975  GAGAGTGAATTTTATGAAAATGGAGCAACATAGCTGAAGATGACAAACATGTCCAC 2034
QY      2219  AAACAACATCTCCACAGAAACCCAGTCTCTGAATTTGGTCCAGTTTGTAGACAACACC 2278
Db      2035  GCACAGCAGTCCCGGTTAGAACCAGGCTCCGGTGGTCTGGCTTTGTACACACCAG 2094
QY      2279  TTTGCTGAAGAAATTCACACTCAGAAATCAGAAATCCAGGATGTGAACTCTGGAGGGA 2338
Db      2095  GCTGCTAAAGAAATTCACACTCAGAAATCAAAATCCAGGATGTGGGTTCTGGAGGGA 2154
QY      2339  GAAGTGGTCAAGAGCTCTCTGTGGAAGACAGATGAAGAGAAATCGGTATTATGATGAG 2398
Db      2155  GAAGTGGTCCGAGAGCTCTGTGGAGAACAGATGAAGAGAAACCGGTACTATGATGAG 2214
QY      2399  GATGAGATGAAGAGTGAACAAATTCGAATGATGTGGCCCTTAAATTCATGTAGTGTTA 2458
Db      2215  GACGAGGATGAAGATGACCGGCCCATGGTTCTGGCC-----TTAGATTTTGGTTA 2267
QY      2459  GCGAGCAGTCCCTTTGTCAAAATGTGATGACATGAAGCAGGTATCCAGCATGAAATG 2518
Db      2268  GCAACTCACTGCCCTTTATCAA---GTGGTCCGAGAAACAGGAGCTTGGCGCCNAGGG 2324
QY      2519  TAAATTTACTTCGAAAGTAACTTTGGAAGAGAAATTCCTTCT-TAAATCAAAACAAACAA 2577
Db      2325  TCATTCATGTGGGAGCAACTTTGGAATGAATTCCTCGTAGCAAAACAGGCAAAACAA 2384
QY      2578  AAAAACACAAAAACACATTTCTAAAT 2603
Db      2385  GCAAAAACAAAACAGACAAACAACT 2410
```

RESULT 3

AK028186

LOCUS

DEFINITION

AK028186 3100 bp mRNA linear HTC 05-DEC-2002
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610034J22 product:epithelial protein lost
in neoplasm, full insert sequence.

ACCESSION

AK028186.1

VERSION

AK028186.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM	Mus musculus	TITLE	Direct Submission
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.go.jp).
AUTHORS	Carninci, P. and Hayashizaki, Y.	COMMENT	URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216
TITLE	High-efficiency full-length cDNA cloning		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		Please visit our web site for further details.
MEDLINE	99279253	URL:	http://genome-gsc.riken.go.jp/
PUBMED	10349636	URL:	http://fantom.gsc.riken.go.jp/
REFERENCE		FEATURES	Location/Qualifiers
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	1. 3100	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	/organism="Mus musculus"	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	/mol_type="mRNA"	
MEDLINE	20493374	/strain="C57BL/6J"	
PUBMED	11042159	/db_xref="FANTOM_DB:2610034J22"	
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kikunishi, Y., Konno, H., Adachi, J., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	/db_xref="taxon:10090"	
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	/clone="2610034J22"	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	/tissue type="whole body"	
MEDLINE	20530913	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
PUBMED	11076861	/dev_stage="10 days embryo"	
REFERENCE		448..2229	
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schrim, L.M., Stauble, F., Suzuki, R., Tonita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M.P., Brownstein, M.J., Bult, C., Fleischer, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.	/note="unnamed protein product; epithelial protein lost in neoplasm (MGD GI:1920992, GB NM_023063, evidence: BLASTN, 99%, match=3086)"	
TITLE	Functional annotation of a full-length mouse cDNA collection	/protein_id="BAC25798.1"	
JOURNAL	Nature 409 (6821), 685-690 (2001)	/db_xref="GI:26389834"	
MEDLINE	21085660	/translation="MENCIGDSRHAERKPESTENTSGKIEKYNVPLNRLKMPFEKSEHNSIKDRMAKYCAAVSVKSSPASVYNELKTESKTHKWEKENVPPGPEACSVHOEGSKVSTENSLVSAVPAEDDTNSQVSEAPQMPKPLSPDARTSSLPESPSTAKKHPENLQFKSGNYDEFGHGHKDLWASKSDNEETLGRPAOPPNAGSPHSPGVEDA PLAKVGVLAASMEAKASOREDEKPAETKLRIAWPPPAELGGGSALEEGIKVSKP KPPEDDVCKTEAPEDVDLKLRSLSLKERSPFVAASFRSSISPKASPSL RKGWSESESEFEFGGIATMERKOTENARPSEKENVKSRWQGEVPRSKDRSPFL ESENFMENAGIAEDDNDVHVAOOSPLEPEAPGSGFVDTAAKEFTTHNQKSDQGVF EGEVRELSVEEQIKENRYDEDEEE"	
PUBMED	11217851	/codon_start=1	
REFERENCE		/base_count 908 a 759 c 824 g 609 t	
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	ORIGIN	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	Query Match 37.8%; Score 1401.6; DB 11; Length 3100;	
JOURNAL	Nature 420, 563-573 (2002)	Best Local Similarity 77.5%; Pred. No. 9.3e-171;	
MEDLINE	6 (bases 1 to 3100)	Matches 1787; Conservative 0; Mismatches 484; Indels 35; Gaps 6;	
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akaira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, F., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	QY 302 AACACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAG 361	
		DB 133 AACCTGAAAGTCTGCCAGCAGCTTTAGAGAGGACCCCTGCTGTGTTAAAGAGAG 192	
		QY 362 TGGGAGAACCCAGGCTGGGAGCAGAGTCTCACAGAGCTCTCTACGGAACAGCAGCACT 421	
		DB 193 TGGGAGAACCCGGTGGCTGGGCGAGAAATTCACAGAGCTCACTCCCAACAGCAGCAGT 252	
		QY 422 GAGATTAGCAGCAGAGAGACCATCTCTCTGCTGAAGTGACAAGCCAGCGCTGCTTCTGGA 481	
		DB 253 GAGGTGGGCGACACAGCGAGCTACCCCTCTGCTGAAGTGACGGACAAGCCCTGCTCTGGA 312	
		QY 482 GCCAAGCTGCCAGAGAGACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGTA 541	
		DB 313 GTCAGAGCTGACCGGGAAGAGACACACCCAGGCCAAACCTAGATTGAGTCTCGTCCGGA 372	
		QY 542 GCCCTCGTTCAGGGTCGATATCCCCACATCAAGACCGTGAGGATCTTAAGACACTCA 601	
		DB 373 GCAGTTATCCAGAGCCCGGTATCTCGCTCAGAGAACGCCACGATTTTAAAGCCCGCC 432	

QY 602 ACAGAAAGTAAATAATGGAATTTGTCTAGAGAAATCCAGGCATGAAGATAGAAAAATCA 661
Db 433 ACGGAGAGCAAAAAATGAAAAATGTCTGGAGATTCAGGCATGAAGCAGAGAAGCCA 492
QY 662 GAAATCAGTGAATAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 721
Db 493 GAGACGAGCGAATAACACAGAACTTCAGGCAAAATAGAGAAATACAAAGTTCCACTGAAT 552
QY 722 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTTAAGATTTCTCCGGSCCCAA 781
Db 553 AGACTGAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTTAAGATTTCTCCGGSCCCAA 612
QY 782 AGCCCAAGTGAAGTGAAGGAAAGATCTCTGAAAACAGCTATTTCTTAGATGACCTGGAA 841
Db 613 AGCCCAAAATGCGGTGGAGAGGCTCTCTGAAAACAACTGTTCCCTCGATGACTGGAA 672
QY 842 ATAGCCCCAGGTCAGTTGCTATCTTACATTTGACTCGGAGAAAATAGAGTAGACA 901
Db 673 ATAGTGGGGTCAATTTGTTCTCTCTGATTCAACTCGAGAAAACAGAGTAAGAGG 732
QY 902 AATCTGGAATTTCCACGGCTCTCGAAAACCTCTATAAAGGATCGAATGGCCCAACTACCAG 961
Db 733 AATCTGGAGTCCCAAGCTCTCGAAAACCTCTATAAAGGATCGAATGGCCCAACTACCAG 792
QY 962 GCAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 1021
Db 793 GCTGCAGTGTCCAAAGCAGCAGCCAGCCAGCTACACAAATGAGCTGAAAACAGT--- 849
QY 1022 GGCGAATCAAAATTCATAAATGAGCAAAAGAGAGATGTGCCCCAGGTCTCTGAGGTC 1081
Db 850 ---GAAAGCAAACTCAATAATGGGAACAGAAAGAGAAATGTGCCCCAGGTCTCCGAGGCC 906
QY 1082 TGCATCACCCATCAGGAAGGGGAAAGATTTCTCCAAATGAGATAGCCTGGCAGTCCGT 1141
Db 907 TGCAGCTCCATCAGGAAGGAAGCAAGGTTTCTACAACTGAGAAATAGTCTAGTGGCCCTC 966
QY 1142 TCCACCCCTCGCGAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAAGTTCAACAGCCT 1201
Db 967 TCTGTCCCTGCTGAAGATGACACCTGTAACTCCAGGTGAAGCAGGAGCCAGAGCCT 1026
QY 1202 GTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCCAGTCTTTCTGAAAGTTCTCCT 1261
Db 1327 ATGACACCTTAAGCCGCTGAGTCTGAGCCAGGAGCTCCAGTCTTCGGAAAGTTCTCCT 1086
QY 1262 CCCAAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTGAGAAG 1321
Db 1087 TCCAAAACAGCAAGAAGTTTCAGGCGCCGCAAGAGAGAGCTGCGTGGAGTGTGAGAAG 1146
QY 1322 ACAGTCTATCCAAATGGAGGCTCTCTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTTC 1381
Db 1147 ACGGTGTACCCCATGGACCGCTCTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTTTC 1206
QY 1382 CGTTGCTCTATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAAGA 1441
Db 1207 CGATGCTCTCTACTGCAACAAACAAAGCTCAGTCTAGGAACATATGCATCTTTACATGGAAGA 1266
QY 1442 ATCTATTGAAGCTCAGTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGC 1501
Db 1267 ATCTATTGAAGCTCAGTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGC 1326
QY 1502 TTTGGGCACAGACACACAGGATCTATGGGCAAGCAAAATGAAACCGAAGAGATTTTG 1561
Db 1327 TTTGGGCACAGCAGCATGAAGATCTGTGGGCAAGCAAGAGCGCAATGAGGAGACTTTG 1386
QY 1562 GAGAGACCGCCAGCTTGCATGCAAGGAGAGCCCTCACAGCCCGAGGCTAGAGAT 1621
Db 1387 GGGAGACCGCCAGCCGCTTATGAGGGGAGAGCCCTCACAGCCCGAGGCTAGAGAT 1446
QY 1622 GCCCTATTGCTAAGTGGGTCTCTGGTCTGAAAGTGAAGCCAGCCCTCTCTCAG 1681
Db 1447 GCCCCCATCGCAAGGTGCGCTGTGGCGCAAGTATGGAAGCCAGGCTCTCTCAG 1506
QY 1682 CAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCCCTC 1741

Db 1507 AGGAGAGGGAGGATTAAGCCTCGGAGACCAAGAGCTGAGGATCGCCTGCCCTCCGCCA 1566
QY 1742 ACTGAACTTGGAGTTTCAAGAAAGTCCCTTGGAGAAAGGATCAAAATGTCAAAGCCCAAA 1801
Db 1567 GCCAGCTGGCGGTTCCGGAAGTGCCTTGGAGAAAGGATCAAAGTATCGAAGCCCAAG 1626
QY 1802 TGGCTTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTG 1861
Db 1627 TGGCTTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTG 1686
QY 1862 AAGAAGCTAAGACGATCTTCTTCACTGAAGAAAGAGCCGCCCATCTCTAGAGAGCT 1921
Db 1687 AAGAAGCTGCGGGGCTTCTTCTTCACTGAAGAAAGAGCCGCCCATCTCTAGAGAGCT 1746
QY 1922 TCATTTCAAGACACCTCTCTGCAAGAGCCCAAAACCTGTGTCCCCACCTATCAGAGAGGC 1981
Db 1747 TCCTTTGAACTCTTCCATCAAGAGCCCAAGGCTCTTTCGCCATCTCTCAGAGAAAGGT 1806
QY 1882 TGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGAAGAGTTGCA---GAAAGGAAA 2038
Db 1807 TCGAGCAGTCCGAGCAGAGTGAAGAGTCTGTGGAGGAGGATAGCGAGGATGGAAGGAAA 1866
QY 2039 CAAGTGGAAAATGCCAAGGCTTCTTAAGAAATGGGAATGTGGAAAACAACTGGCAA 2098
Db 1867 CAAACCGAAAATGCCAAGGCTTCTGTGGGAGAAAGAGAAATGTGGAAAATCACCCCTGGCAG 1926
QY 2099 AACAAAGAAATCTAAAGAGAGACAGAGAGAGAAAGTAAAGAAAGTCAAGTTTGGAGATG 2158
Db 1927 -----GGCAGGAAGTGCAGAGAGCAAGGACCGTAGTTCTTTTGAGCTG 1971
QY 2159 GAGAAATGAGAAATCTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCCTC 2218
Db 1972 GAGAGTGAAGATTTTATGAAAATGGAGCAACATAGCTGAAGATGACAACTATGTCAC 2031
QY 2219 AAACAACAATCTCCACAGAACCCAGTCTCTGAATTTGGTTCGAGTTTGTAGACAACACC 2278
Db 2032 GCACAGCAGTCCCGGTAGAACCCGAGGCTCCGGTTCGGTCTGGCTTTGTAGACACCAG 2091
QY 2279 TTTGCTGAAGATTTCACTACTCAGAAATCAGAAATCCCAGGATGTGGAATCTCGGAGGGA 2338
Db 2092 GTGCTTAAAGAAATTCACACCCCAATCAAAATCCCAAGATGTCCGGTCTTCGGAGGGA 2151
QY 2339 GAAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATTAAGAGAAATCCGTTATATGATGAG 2398
Db 2152 GAAGTGGTCCGAGAGCTGTCTGTGGAGAACAGATTAAGAGAAACCGGTACTATGATGAG 2211
QY 2399 GATGAGGATGAAGAGTGACAAATTCGAATGATGTGGGCTTAAATTCATGTTAGTGTTA 2458
Db 2212 GACGAGGATGAAGATGACCGGCCCGCATGGTCTGGCC-----TTAGATTTTGGTTA 2264
QY 2459 GCGAGCAGTCCCTTTGTCAAAATGTGATCAGATAGCAGGATCCAGCATGAAATG 2518
Db 2265 GCAACTCAGTCCCTTTATCAA---GTGGTGGCAGAAACAGGAGCTTGGGCCAAGG 2321
QY 2519 TAATTTACTTGAAGTAACTTTGAAAAGAAATTCCTCT---TAAATCAAAAACAAAAA 2577
Db 2322 TCATTCATGTGGAGCAACTTTGAAAATGAATTCCTGCTAGCAAAACAGGCAAAACAA 2381
QY 2578 AAAAAACAAAAACACATTTCTAAAT 2603
Db 2382 GCAAAAACAAAAACAGACAAACAACT 2407

RESULT 4
AK085065
LOCUS
DEFINITION
Mus musculus 13 days embryo lung cDNA, Riken full-length enriched
library, clone:D430032103 product:epithelial protein lost in
neoplasm, full insert sequence.
ACCESSION
AK085065
VERSION
AK085065.1 GI:26351432
KEYWORDS
HTC; CAP trapper.

AK085065 3160 bp mRNA linear HTC 05-DEC-2002

